520 Rec'd PCT/PTO 01 NOV 1999

PCT

FORM PTO-1390 (REV 1-98) U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

ATTORNEY'S DOCKET NUMBER: 23541-01

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

U.S. APPLICATION NO. (If known, see 37 CFR 1.5): 09/425095

INTERNATIONAL APPLICATION NO	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/AU98/00315	01/05/1998 (1 MAY 1998)	01/05/1997 (01 MAY 1997)

TITLE OF INVENTION: NUCLEIC ACID MOLECULES SPECIFIC FOR BACTERIAL ANTIGENS AND USES THEREOF

APPLICANT(S) FOR DO/EO/US: Peter Richard REEVES and Lei WANG

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. [X] This is a FIRST submission of items concerning a filing under 35 U.S.C. 371.
2. [ ] This is a SECOND or SUBSEQUENT submission of items concerning a filing under 35 U.S.C. 371.
3. [X] This express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).
4. [X] A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. [X] A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. [X] is transmitted herewith (required only if not transmitted by the International Bureau).
   b. [ ] has been transmitted by the International Bureau.
   c. [ ] is not required, as the application was filed in the United States Receiving Office (RO/US)
6. [ ] A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. [X] Amendments to the claims of the International Aplication under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. [ ] are transmitted herewith (required only if not transmitted by the International Bureau).
   b. [ ] have been transmitted by the International Bureau
   c. [ ] have not been made; however, the time limit for making such amendments has NOT expired.
   d. [X] have not been made and will not be made.
8. [ ] A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).
9. [X] An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. [ ] A translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11. to 16. below concern document(s) or information included:

11. [ ] An Information Disclosure Statement under 37 CFR 1.97 and 1.98
12. [ ] An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. [X] A FIRST preliminary amendment.
    [ ] A SECOND or SUBSEQUENT preliminary amendment.
14. [ ] A substitute specification.
15. [ ] A change of power of attorney and/or address letter.
16. [X] Other items or information:

- PCT Publication No. WO 98/50531
- PCT Request
- PCT Chapter II Demand
- International Search Report and Citations
- International Preliminary Examination Report
- Written Opinion

EXPRESS MAIL Label No. EL007669381US - November 1, 1999

U.S. APPLICATION NO. (IF KNOWN, SEE 37 CFR 1.5) 09/423093	INTERNATIONAL APPLICATION NO PCT/AU98/00315	ATTORNEY'S DOCKET NUMBER 23541-01

17. ☒ The following fees are submitted:

	CALCULATIONS PTO USE ONLY	
BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):		
Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO $1070.00		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO $930.00		
International preliminary examination fee (37 CFR 1 482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO $790.00		
International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) $720.00		
International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) $98.00		
ENTER APPROPRIATE BASIC FEE AMOUNT =	$ 1070.00	
Surcharge of $130 00 for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1 492(e)).	$	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	$	
Total claims	52 - 20 =	32	x $22.00	$ 576	
Independent claims	16 - 3 =	13	x $82 00	$ 1014	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)		1	+ $270.00	$ 260	
TOTAL OF ABOVE CALCULATIONS =				$ 2920	
Reduction of 1/2 for filing by small entity, if applicable. A Small Entity Statement must also be filed (Note 37 CFR 1.9, 1.27, 1.28). +				$ -1460	
SUBTOTAL =				$ 1460	
Processing fee of $130 00 for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(f)).				$	
TOTAL NATIONAL FEE =				$ 1460	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). $40.00 per property +				$	
TOTAL FEES ENCLOSED =				$ 1460	
				Amount to be refunded:	$
				charged:	$

a. ☒ A check in the amount of $ 1460 to cover the above fees is enclosed

b ☐ Please charge my Deposit Account No. _____ in the amount of $ _____ to cover the above fees. A duplicate copy of this sheet is enclosed.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No 03-3415 . A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

William H. Dippert
Cowan Liebowitz & Latman, P.C.
1133 Avenue of the Americas
New York, NY 10036-6799

SIGNATURE

William H. Dippert
NAME

26.723
REGISTRATION NUMBER

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Peter Richard REEVES, et al.

Serial No.: to be assigned

Filed: to be assigned

For: NUCLEIC ACID MOLECULES SPECIFIC FOR BACTERIAL ANTIGENS AND USES THEREOF

November 1, 1999

Asst. Commissioner for Patents
U.S. Patent and Trademark Office
Washington, D.C. 20231

PRELIMINARY AMENDMENT

S I R :

Prior to examination or calculation of the filing fee, please amend the above-referenced application as follows:

IN THE CLAIMS:

At lines 1 and 2 of each of Claims 27 and 28, change "any one of claims 22 to 26" to -- claim 22 --.

Claim 29, lines 3 and 4, Claim 30, lines 3 and 4, and Claim 31, lines 3 and 4, change "any one of claims 16 to 28" to -- claim 16 or 28 --.

Claim 42, line 1, change "31" to -- 32 --.

Respectfully submitted,

William H. Dippert
Reg. No. 26,723

COWAN, LIEBOWITZ & LATMAN, P.C.
1133 Avenue of the Americas
New York, NY 10036-6799
(212) 790-9200

EXPRESS MAIL CERTIFICATE 37 C.F.R. 1.10

Date of Deposit November 1, 1999

Express Mail Label No. EL007669381US

I hereby certify that this paper is being deposited with the U.S. Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to Commissioner of Patents and Trademarks, Washington, D.C. 20231

Eugene Acevedo
Name of Person Mailing

Signature

PTO/SB/11 (12-97)
Approved for use through 9/30/00. OMB 0651-0031
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT CLAIMING SMALL ENTITY STATUS (37 CFR 1.9(f) & 1.27(d))--NONPROFIT ORGANIZATION

Docket Number (Optional): 23541-01

Applicant, Patentee, or Identifier: Peter Richard REEVES and Lei WANG
Application or Patent No.: to be assigned
Filed or Issued: to be assigned
Title: NUCLEIC ACID MOLECLES SPECIFIC FOR BACTERIAL ANTIGENS AND USES ...

I hereby state that I am an official empowered to act on behalf of the nonprofit organization identified below:
NAME OF NONPROFIT ORGANIZATION The University of Sydney
ADDRESS OF NONPROFIT ORGANIZATION Parramatta Road, Sydney NSW, Australia 2006

TYPE OF NONPROFIT ORGANIZATION:

☒ UNIVERSITY OR OTHER INSTITUTION OF HIGHER EDUCATION

☐ TAX EXEMPT UNDER INTERNAL REVENUE SERVICE CODE (26 U.S.C. 501(a) and 501(c)(3))

☐ NONPROFIT SCIENTIFIC OR EDUCATIONAL UNDER STATUTE OF STATE OF THE UNITED STATES OF AMERICA
(NAME OF STATE _____)
(CITATION OF STATUTE _____)

☐ WOULD QUALIFY AS TAX EXEMPT UNDER INTERNAL REVENUE SERVICE CODE (26 U.S.C. 501(a) and 501(c)(3)) IF LOCATED IN THE UNITED STATES OF AMERICA

☐ WOULD QUALIFY AS NONPROFIT SCIENTIFIC OR EDUCATIONAL UNDER STATUTE OF STATE OF THE UNITED STATES OF AMERICA IF LOCATED IN THE UNITED STATES OF AMERICA
(NAME OF STATE _____)
(CITATION OF STATUTE _____)

I hereby state that the nonprofit organization identified above qualifies as a nonprofit organization as defined in 37 CFR 1.9(e) for purposes of paying reduced fees to the United States Patent and Trademark Office regarding the invention described in:

☒ the specification filed herewith with title as listed above.
☐ the application identified above.
☐ the patent identified above.

I hereby state that rights under contract or law have been conveyed to and remain with the nonprofit organization regarding the above identified invention. If the rights held by the nonprofit organization are not exclusive, each individual, concern, or organization having rights in the invention must file separate statements as to their status as small entities and that no rights to the invention are held by any person, other than the inventor, who would not qualify as an independent inventor under 37CFR 1.9(c) if that person made the invention, or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d) or a nonprofit organization under 37 CFR 1.9(e).

Each person, concern, or organization having any rights in the invention is listed below:

☒ no such person, concern, or organization exists.
☐ each such person, concern, or organization is listed below.

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 CFR 1.28(b))

NAME OF PERSON SIGNING X Claire Baxter CLAIRE-BAXTER
TITLE IN ORGANIZATION OF PERSON SIGNING X DIRECTOR BUSINESS LIAISON OFFICE
ADDRESS OF PERSON SIGNING X UNIVERSITY OF SYDNEY, AUSTRALIA 2006
SIGNATURE X Claire Baxter DATE X OCT 19 1999

Nucleic acid molecules specific for bacterial antigens and uses thereof.

TECHNICAL FIELD

The invention relates to novel nucleotide sequences located in a gene cluster which controls the synthesis of a bacterial polysaccharide antigen, especially an O antigen, and the use of those nucleotide sequences for the detection of bacteria which express particular polysaccharide antigens (particularly O antigens) and for the identification of the polysaccharide antigens (particularly O antigens) of those bacteria.

BACKGROUND ART

Enteropathogenic E. coli strains are well known causes of diarrhoea and haemorrhagic colitis in humans and can lead to potentially life threatening sequelae including haemolytic uremic syndrome and thrombotic thrombocytopaenic purpura. Some of these strains are commonly found in livestock and infection in humans is usually a consequence of consumption of contaminated meat or dairy products which have been improperly processed. The O specific polysaccharide component (the "O antigen") of lipopolysaccharide is known to be a major virulence factor of enteropathogenic E. coli strains.

The E. coli O antigen is highly polymorphic and 166 different forms of the antigen have been defined; Ewing, W. H. [in Edwards and Ewings "Identification of the Enterobacteriacea" Elsevier. Amsterdam (1986)] discusses 128 different O antigens while Lior H. (1994) extends the number to 166 [in "Classification of *Escherichia coli* In *Escherichia coli in domestic animals and humans* pp31-72. Edited by C.L.Gyles CAB International]. The genus Salmonella enterica has 46 known O antigen types [Popoff M.Y. *et al* (1992) " Antigenic formulas of the Salmonella enterica serovars" 6th revision WHO Collaborating Centre for Reference and Research on Salmonella enterica, Institut Pasteur Paris France].

An important step in determining the biosynthesis of O antigens and therefore the mechanism of the polymorphism has been to characterise the gene clusters controlling O antigen biosynthesis. The genes specific for the synthesis of the O antigen are generally located in a gene cluster at map position 45 minutes on the chromosome of E. coli K-12 [Bachmann, B. J. 1990 "Linkage map of Escherichia coli K-12". *Microbiol. Rev.* 54: 130-197], and at the corresponding position in S. enterica LT2 [Sanderson et al (1995) "Genetic map of Salmonella enterica typhimurium", Edition VIII Microbiol. Rev. **59:** 241-303]. In both cases the O antigen gene cluster is close to the *gnd* gene as is the case in other strains of E. coli and S. enterica [Reeves P.R. (1994) "Biosynthesis and assemby of lipopolysaccharide, 281-314. in A. Neuberger and L.L.M. van Deenen (eds) "Bacterial cell wall, new comprehensive biochemistry " vol 27 Elsevier Science Publishers]. These genes encode enzymes for the synthesis of nucleotide diphosphate sugars and for assembly of the sugars into oligosaccharide units and in general for polymerisation to O antigen.

The E. coli O antigen gene clusters for a wide range of E. coli O antigens have been cloned but the O7, O9, O16 and O111 O antigens have been studied in more detail with only O9 and O16 having been fully characterised with regard to nucleotide sequence to date [Kido N., Torgov V.I., Sugiyama T., Uchiya K., Sugihara H., Komatsu T., Kato N. & Jann K. (1995) "Expression of the O9 polysaccharide of *Escherichia coli*: sequencing of the *E. coli* O9 rfb gene cluster, characterisation of mannosyl transferases, and evidence for an ATP-binding cassette transport system" J. of Bacteriol. 177 2178-2187; Stevenson G., Neal B., Liu D., Hobbs M., Packer N.H., Batley M., Redmond J.W., Lindquist L. & Reeves PR (1994) "Structure of the O antigen of *E. coli* K12 and the sequence of its *rfb* gene cluster" J. of Bacteriol. **176** 4144-4156; Jayaratne, P. *et al.* (1991) "Cloning and analysis of duplicated *rfbM* and *rfbK* genes involved in the

formation of GDP-mannose in *Escherichia coli* O9:K30 and participation of *rfb* genes in the synthesis of the group 1 K30 capsular polysaccharide" *J. Bacteriol.* 176: 3126-3139; Valvano, M. A. and Crosa, J. H.(1989)" Molecular cloning and expression in *Escherichia coli K-12* of chromosomal genes determining the O7 lipopolysaccharide antigen of a human invasive strain of *E.coli* O7:K1". *Inf and Immun.* 57:937-943; Marolda C. L. And Valvano, M. A. (1993). "Identification, expression, and DNA sequence of the GDP-mannose biosynthesis genes encoded by the O7 *rfb* gene cluster of strain VW187 (*Eschericia coli* O7:K1)". *J. Bacteriol.* 175:148-158.]

Bastin D.A., *et al.* 1991 ["Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E.coli* O111 strain". *Mol. Microbiol.* 5:9 2223-2231] and Bastin D.A. and Reeves, P.R. [(1995)" Sequence and analysis of the O antigen gene(*rfb*)cluster of *Escherichia coli* O111". *Gene* 164: 17-23] isolated chromosomal DNA encoding the E. coli O111 *rfb* region and characterised a 6962 bp fragment of E. coli O111 rfb. Six open reading frames (orfs) were identified in the 6962 bp partial fragment and the alignment of the sequences of these orfs revealed homology with genes of the GDP-mannose pathway, *rfbK* and *rfbM*, and other *rfb* and *cps* genes.

The nucleotide sequences of the loci which control expression of Salmonella enterica B, A, D1, D2, D3, C1, C2 and E O antigens have been characterised [Brown, P. K., L. K. Romana and P. R. Reeves (1991) "Cloning of the *rfb* gene cluster of a group C2 Salmonella enterica: comparison with the *rfb* regions of groups B and D Mol. Microbiol. 5:1873-1881; Jiang, X.-M., B. Neal, F. Santiago, S. J. Lee, L. K. Romana, and P. R. Reeves (1991) "Structure and sequence of the *rfb* (O antigen) gene cluster of Salmonella enterica serovar typhimurium (LT2)". Mol. Microbiol. 5:692-713; Lee, S. J., L. K. Romana, and P. R. Reeves (1992) "Sequences and structural analysis of the *rfb* (O antigen)gene cluster from a group C1 Salmonella enterica

enterica strain" J. Gen. Microbiol. **138**: 1843-1855; Lui, D., N. K. Verma, L. K. Romana, and P. R. Reeves (1991) "Relationship among the *rfb* regions of Salmonella enterica serovars A, B and D" J. Bacteriol. **173:** 4814-4819; Verma, N. K., and P. Reeves (1989) "Identification and sequence of *rfb*S and *rfb*E, which determine the antigenic specificity of group A and group D Salmonella entericae" J. Bacteriol. **171:** 5694-5701; Wang, L., L. K. Romana, and P. R. Reeves (1992) "Molecular analysis of a Salmonella enterica enterica group E1 *rfb* gene cluster: O antigen and the genetic basis of the major polymorphism" Genetics **130:** 429-443; Wyk, P., and P. Reeves (1989). "Identification and sequence of the gene for abequose synthase, which confers antigenic specificity on group B Salmonella entericae: homology with galactose epimerase" J. Bacteriol. **171:** 5687-5693,; Xiang, S. H., M. Hobbs, and P. R. Reeves. 1994 Molecular analysis of the rfb gene luster of a group D2 Salmonella enterica strain: evidence for its origin from an insertion sequence -mediated recombination event between group E and D1 strains. J. Bacteriol. 176: 4357 -4365; Curd, H., D. Liu and P. R. Reeves, 1998. Relationships among the O antigen Salmonella enterica groups B, D1, D2, and D3. J. Bacteriol. 180: 1002-1007.].

Of the closely related Shigella (which really can be considered to be part of E. coli) S. dysenteriae and S. flexneri O antigens have been fully sequenced and are next to *gnd*. [Klena JD & Schnaitman CA (1993) "Function of the *rfb* gene cluster and the *rfe* gene in the synthesis of O antigen by *Shigella dysenteriae* 1" Mol. Microbiol. **9** 393-402; Morona R., Mavris M., Fallarino A. & Manning P. (1994) "Characterisation of the *rfc* region of *Shigella flexneri*" J.Bacteriol **176**: 733-747]

Inasmuch as the O antigen of enteropathogenic E. coli strains and the O antigen of Salmonella enterica strains are major virulence factors and are highly polymorphic, there is a real need to develop highly specific, sensitive, rapid and inexpensive diagnostic assays to

detect E. coli and assays to detect S. enterica. There is also a real need to develop diagnostic assays to identify the O antigens of E. coli strains and assays to identify the O antigens of S. enterica strains. With regard to the detection of E. coli these needs extend beyond EHEC (enteropathogenic haemorrhagic E. coli) strains but this is the area of greatest need. There is interest in diagnostics for ETEC (enterotoxigenic E. coli) etc in E. coli.

The first diagnostic systems employed in this field used large panels of antisera raised against E. coli O antigen expressing strains or S. enterica O antigen expressing strains. This technology has inherent difficulties associated with the preparation, storage and usage of the reagents, as well as the time required to achieve a meaningful diagnostic result.

Nucleotide sequences derived from the O antigen gene clusters of S. enterica strains have been used to determine S. enterica O antigens in a PCR assay [Luk, J.M.C. *et al.* (1993) "Selective amplification of abequose and paratose synthase genes (*rfb*) by polymerase chain reaction for identification of S. enterica major serogoups (A, B, C2, and D)", *J. Clin. Microbiol.* 31:2118-2123]. The prior complete nucleotide sequence characterisation of the entire *rfb* locus of serovars Typhimurium, Paratyphi A, Typhi, Muenchen, and Anatum; representing groups B, A, D1, C2 and E1 respectively enabled Luk *et al.* to select oligonucleotide primers specific for those serogroups. Thus the approach of Luk *et al.* was based on aligning known nucleotide sequences corresponding to CDP-abequose and CDP-paratose synthesis genes within the O antigen regions of S. enterica serogroups E1, D1, A, B and C2 and exploiting the observed nucleotide sequence differences in order to identify serotype-specific oligonucleotides.

In an attempt to determine the O antigen serotype of a Shiga-like toxin producing E. coli strain, Paton, A. W., *et al.* 1996 ["Molecular microbiological investigation of an outbreak of Hemolytic-Uremic Syndrome caused by dry

fermented sausage contaminated with Shiga-like toxin producing Escherichia coli". *J. Clin. Microbiol.* **34:** 1622-1627], used oligonucleotides derived from the *wbdI* (*orf6*) region, which were believed to be specific to the E. coli O111 antigen and which were derived from E. coli O111 sequence, in a PCR diagnostic assay. Unpublished reports indicate that the approach of Paton *et al.* is deficient in that the nucleotide sequences derived from *wbdI* may not specifically identify the O111 antigen and in fact lead to detection of false positive results. Paton *et al.* disclose the detection of 5 O111 antigen isolates by PCR when in fact from only 3 of those isolates did they detect bacteria which reacted with O111 specific antiserum.

DESCRIPTION OF THE INVENTION

Whilst not wanting to be held to a particular hypothesis, the present inventors now believe that the reported false positives found with the Paton *et al.* method are due to the fact that the nucleic acid molecules employed by Paton *et al.* were derived from genes which have a putative function as a sugar pathway gene, [Bastin D.A. and Reeves, P.R. (1995) Sequence and analysis of the O antigen gene(*rfb*) cluster of Escherichia coli O111. *Gene* 164: 17-23] which they now believe to lack the necessary nucleotide sequence specificity to identify the E. coli O antigen. The inventors now believe that many of the nucleic acid molecules derived from sugar pathway genes expressed in S. enterica or other enterobacteria are also likely to lack the necessary nucleotide sequence specificity to identify specific O antigens or specific serotypes.

In this regard it is important to note that the genes for the synthesis of a polysaccharide antigen include those related to the synthesis of the sugars present in the antigen (sugar pathway genes) and those related to the manipulation of those sugars to form the polysaccharide. The present invention is predominantly concerned with the latter group of genes, particularly the assembly and

transport genes such as transferase, polymerase and flippase genes.

The present inventors have surprisingly found that the use of nucleic acid molecules derived from particular assembly and transport genes, particularly transferase, wzx and wzy genes, within O antigen gene clusters can improve the specificity of the detection and identification of O antigens. The present inventors believe that the invention is not necessarily limited to the detection of the particular O antigens which are encoded by the nucleic acid molecules exemplified herein, but has broad application for the detection of bacteria which express an O antigen and the identification of O antigens in general. Further because of the similarities between the gene clusters involved in the synthesis of O antigens and other polymorphic polysaccharide antigens, such as bacterial capsular antigens, the inventors believe that the methods and molecules of the present invention are also applicable to these other polysaccharide antigens.

Accordingly, in one aspect the present invention relates to the identification of nucleic acid molecules which are useful for the detection and identification of specific bacterial polysaccharide antigens.

The invention provides a nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, including a *wzx* gene, *wzy* gene, or a gene with a similar function; the gene being involved in the synthesis of a particular bacterial polysaccharide antigen,
wherein the sequence of the nucleic acid molecule is specific to the particular bacterial polysaccharide antigen.

Polysaccharide antigens, such as capsular antigens of E. coli (Type I and Type II), the Virulence capsule of S. enterica sv Typhi and the capsules of species such as Streptococcus pneumoniae and Staphylococcus albus are

encoded by genes which include nucleotide sugar pathway genes, sugar transferase genes and genes for the transport and processing of the polysaccharide or oligosaccharide unit. In some cases these are *wzx* or *wzy* but in other cases they are quite different because a different processing pathway is used. Examples of other gene clusters include the gene clusters for an extracellular polysaccharide of Streptococcus thermophilus, an exopolysaccharide of Rhizobium melilotti and the K2 capsule of Klebsiella pneumoniae. These all have genes which by experimental analysis, comparison of nucleotide sequence or predicted protein structure, can be seen to include nucleotide sugar pathway genes, sugar transferase genes and genes for oligosaccharide or polysaccharide processing.

In the case of the E. coli K-12 colanic acid capsule gene cluster [Stevenson *et al* (1996) "Organization of the *Escherichia coli K-12* gene cluster responsible for production of the extracellular polysaccharide colanic acid". J. Bacteriol **178**: 4885-4893] genes from the three classes were identified either provisionally or definitively. Colanic acid capsule is classified with the Type I capsule of E. coli.

The present inventors believe that, in general, transferase genes and genes for oligosaccharide processing will be more specific for a given capsule than the genes coding for the nucleotide sugar synthetic pathways as most sugars present in such capsules occur in the capsules of different serotypes. Thus the nucleotide sugar synthesis pathway genes could now be predicted to be common to more than one capsule type.

As elaborated below the present inventors recognise that there may be polysaccharide antigen gene clusters which share transferase genes and/or genes for oligosaccharide or polysaccharide processing so that completely random selection of nucleotide sequences from within these genes may still lead to cross-reaction; an example with respect to capsular antigens is provided by

the E. coli type II capsules for which only transferase genes are sufficiently specific. However, the present inventors in light of their current results nonetheless consider the transferase genes or genes controlling oligosaccharide or polysaccharide processing to be superior targets for nucleotide sequence selection for the specific detection and characterisation of polysaccharide antigen types. Thus where there is similarity between particular genes, selection of nucleotide sequences from within other transferase genes or genes for oligosaccharide or polysaccharide processing from within the relevant gene cluster will still provide specificity, or alternatively the use of combinations of nucleotide sequences will provide the desired specificity. The combinations of nucleotide sequences may include nucleotide sequences derived from pathway genes together with nucleotide sequences derived from transferase, *wzx* or *wzy* genes.

Thus the invention also provides a panel of nucleic acid molecules wherein the nucleic acid molecules are derived from a combination of genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes; wherein the combination of genes is specific to the synthesis of a particular bacterial polysaccharide antigen and wherein the panel of nucleic acid molecules is specific to a bacterial polysaccharide antigen. In another preferred form, the nucleic acid molecules are derived from a combination of genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, together with nucleic acid molecules derived from pathway genes.

In a second aspect the present invention relates to the identification of nucleic acid molecules which are useful for the detection of bacteria which express O antigens and for the identification of the O antigens of those bacteria in diagnostic assays.

The invention provides a nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit such as a *wzx* or *wzy* gene, the gene being involved in the synthesis of a particular bacterial O antigen, wherein the sequence of the nucleic acid molecule is specific to the particular bacterial O antigen.

The nucleic acids of the invention may be variable in length. In one embodiment they are from about 10 to about 20 nucleotides in length.

In one preferred embodiment, the invention provides a nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit including a *wzx* or *wzy* gene the gene being involved in the synthesis of an O antigen expressed by E. coli, wherein the sequence of the nucleic acid molecule is specific to the O antigen.

In one more preferred embodiment, the sequence of the nucleic acid molecule is specific to the nucleotide sequence encoding the O111 antigen (SEQ ID NO:1). More preferably, the sequence is derived from a gene selected from the group consisting of *wbdH* (nucleotide position 739 to 1932 of SEQ ID NO:1), *wzx* (nucleotide position 8646 to 9911 of SEQ ID NO:1), *wzy* (nucleotide position 9901 to 10953 of SEQ ID NO:1), *wbdM* (nucleotide position 11821 to 12945 of SEQ ID NO:1) and fragments of those molecules of at least 10-12 nucleotides in length. Particularly preferred nucleic acid molecules are those set out in Table 5 and 5A, with respect to the above mentioned genes.

In another more preferred embodiment, the sequence of the nucleic acid molecule is specific to the nucleotide sequence encoding the O157 antigen (SEQ ID NO:2). More preferably the sequence is derived from a gene selected from the group consisting of *wbdN* (nucleotide position 79 to 861 of SEQ ID NO:2), *wbdO*, (nucleotide position 2011 to 2757 of SEQ ID NO:2), *wbdP* (nucleotide position 5257 to

6471 of SEQ ID NO:2)), *wbdR* (13156 to 13821 of SEQ ID NO:2), *wzx* (nucleotide position 2744 to 4135 of SEQ ID NO:2) and *wzy* (nucleotide position 858 to 2042 of SEQ ID NO:2). Particularly preferred nucleic acid molecules are those set out in Table 6 and 6A.

The invention also provides in a further preferred embodiment a nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit including a *wzx* or *wzy* gene; the gene being involved in the synthesis of an O antigen expressed by Salmonella enterica, wherein the sequence of the nucleic acid molecule is specific to the O antigen.

In one more preferred form of this embodiment, the sequence of the nucleic acid molecule is specific to the nucleotide sequence encoding the S. enterica C2 antigen (SEQ ID NO:3). More preferably the sequence of the nucleic acid molecule is derived from a gene selected from the group consisting of *wbaR* (nucleotide position 2352 to 3314 of SEQ ID NO:3), *wbaL* (nucleotide position 3361 to 3875 of SEQ ID NO:3), *wbaQ* (nucleotide position 3977 to 5020 of SEQ ID NO:3), *wbaW* (nucleotide position 6313 to 7323 of SEQ ID NO:3), *wbaZ* (nucleotide position 7310 to 8467 of SEQ ID NO:3), *wzx* (nucleotide position 1019 to 2359 of SEQ ID NO:3)and *wzy* (nucleotide position 5114 to 6313 of SEQ ID NO:3). Particularly preferred nucleic acid molecules are those set out in Table 7.

In another more preferred form of this embodiment, the sequence of the nucleic acid molecule is specific to the nucleotide sequence encoding the S. enterica B antigen (SEQ ID NO:4). More preferably the sequence is derived from *wzx* (nucleotide position 12762 to 14054 of SEQ ID NO:4) or *wbaV* (nucleotide position 14059 to 15060 of SEQ ID NO:4). Particularly preferred nucleic acid molecules are those set out in Table 8 which are derived from *wzx* and *wbaV* genes.

In a further more preferred form of this embodiment, the sequence of the nucleic acid molecule is specific to

the S. enterica D3 O antigen and is derived from the *wzy* gene.

In yet a further preferred form of this embodiment, the sequence of the nucleic acid molecule is specific to the S. enterica E1 O antigen and is derived from the *wzx* gene.

While transferase genes, or genes coding for the transport or processing of a polysaccharide or oligosaccharide unit, such as a *wzx* or *wzy* gene, are superior targets for specific detection of individual O antigen types there may well be individual genes or parts of them within this group that can be demonstrated to be the same or closely related between different O antigen types such that cross-reactions can occur. Cross reactions should be avoided by the selection of a different target within the group or the use of multiple targets within the group.

Further, it is recognised that there are cases where O antigen gene clusters have arisen from recombination of at least two strains such that the unique O antigen type is provided by a combination of gene products shared with at least two other O antigen types. The recognised example of this phenomenon is the S. enterica O antigen serotype D2 which has genes from D1 and E1 but none unique to D2. In these circumstances the detection of the O antigen type can still be achieved in accordance with the invention, but requires the use of a combination of nucleic acid molecules to detect a specific combination of genes that exists only in that particular O antigen gene cluster.

Thus, the invention also provides a panel of nucleic acid molecules wherein the nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is specific to a bacterial O antigen. Preferably the particular bacterial O antigen is expressed by S. enterica. More preferably,

the panel of nucleic acid molecules is specific to the D2 O antigen and is derived from the E1 *wzy* gene and the D1 *wzx* gene.

The combinations of nucleotide sequences may include nucleotide sequences derived from pathway genes, together with nucleotide sequences derived from transferase, *wzx* or *wzy* genes.

Thus, the invention also provides a panel of nucleic acid molecules, wherein the nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, and sugar pathway genes, wherein the panel of nucleic acid molecules is specific to a particular bacterial O antigen. Preferably the O antigen is expressed S. enterica.

Further it is recognised that there may be instances where spurious hybridisation will arise through initial selection of a sequence found in many different genes but this is typically recognisable by, for instance, comparison of band sizes against controls in PCR gels, and an alternative sequence can be selected.

The present inventors believe that based on the teachings of the present invention and available information concerning polysaccharide antigen gene clusters (including O antigen gene clusters), and through use of experimental analysis, comparison of nucleic acid sequences or predicted protein structures, nucleic acid molecules in accordance with the invention can be readily derived for any particular polysaccharide antigen of interest. Suitable bacterial strains can typically be acquired commercially from depositary institutions.

As mentioned above there are currently 166 defined E. coli O antigens while the S. enterica has 46 known O antigen types [Popoff M.Y. *et al* (1992) "Antigenic formulas of the Salmonella serovars" 6th revision WHO Collaborating centre for Reference and Research on Salmonella, Institut Pasteur Paris France]. Many other genera of bacteria are known to have O antigens and these

include Citrobacter, Shigella, Yersinia, Plesiomonas, Vibrio and Proteus.

Samples of the 166 different E. coli O antigen serotypes are available from Statens Serum Institut, Copenhagen, Denmark.

The 46 S. enterica serotypes are available from Institute of Medical and Veterinary Science, Adelaide, Australia.

In another aspect, the invention relates to a method of testing a sample for the presence of one or more bacterial polysaccharide antigens comprising contacting the sample with at least one oligonucleotide molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing of oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the bacterial polysaccharide antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial polysaccharide antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

Where a single specific oligonucleotide molecule is unavailable a combination of molecules hybridising specifically to the target region may be used. Thus the invention provides a panel of nucleic acid molecules for use in the method of testing of the invention, wherein the nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is specific to a particular bacterial polysaccharide. The panel of nucleic acid molecules can include nucleic acid molecules derived from sugar pathway genes where necessary.

In another aspect, the invention relates to a method of testing a sample for the presence of one or more

bacterial polysaccharide antigens comprising contacting the sample with at least one pair of oligonucleotide molecules, with at least one oligonucleotide molecule of the pair capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the bacterial polysaccharide antigen; under conditions suitable to permit the at least one oligonucleotide molecule of the pair of molecules to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial polysaccharide antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

The pair of oligonucleotide molecules may both hybridise to the same gene or to different genes. Only one oligonucleotide molecule of the pair need hybridise specifically to sequence specific for the particular antigen type. The other molecule can hybridise to a non-specific region.

Where the particular polysaccharide antigen gene cluster has arisen through recombination, the at least one pair of oligonucleotide molecules may be selected to be capable of hybridising to a specific combination of genes in the cluster specific to that polysaccharide antigen, or multiple pairs may be selected to provide hybridisation to the specific combination of genes. Even where all the genes in a particular cluster are unique, the method may be carried out using nucleotide molecules which recognise a combination of genes within the cluster.

Thus the invention provides a panel containing pairs of nucleic acid molecules for use in the method of testing of the invention, wherein the pairs of nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is

specific to a particular bacterial polysaccharide antigen. The panel of nucleic acid molecules can include pairs of nucleic acid molecules derived from sugar pathway genes where necessary.

In another aspect, the invention relates to a method of testing a sample for the presence of one or more particular bacterial O antigens comprising contacting the sample with at least one oligonucleotide molecule capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the particular O antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules. Preferably the bacteria are E. coli or S. enterica. More preferably, the E. coli express the 0157 serotype or the 0111 serotype. More preferably the S. enterica express the C2 or B serotype. Preferably, the method is a Southern blot method. More preferably, the nucleic acid molecule is labelled and hybridisation of the nucleic acid molecule is detected by autoradiography or detection of fluorescence.

The inventors envisage circumstances where a single specific oligonucleotide molecule is unavailable. In these circumstances a combination of molecules hybridising specifically to the target region may be used. Thus the invention provides a panel of nucleic acid molecules for use in the method of testing of the invention, wherein the nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is specific to a particular bacterial O antigen. Preferably the particular bacterial O antigen is

expressed by S. enterica. The panel of nucleic acid molecules can include nucleic acid molecules derived from sugar pathway genes where necessary.

In another aspect, the invention relates to a method of testing a sample for the presence of one or more particular bacterial O antigens comprising contacting the sample with at least one pair of oligonucleotide molecules with at least one oligonucleotide molecule of the pair being capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the particular O antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

Preferably the bacteria are E. coli or S. enterica. More preferably, the E. coli are of the 0111 or the 0157 serotype. More preferably the S. enterica express the C2 or B serotype. Preferably, the method is a polymerase chain reaction method. More preferably the oligonucleotide molecules for use in the method of the invention are labelled. Even more preferably the hybridised oligonucleotide molecules are detected by electrophoresis. Preferred oligonucleotides for use with 0111 which provide for specific detection of 0111 are illustrated in Table 5 and 5A with respect to the genes *wbdH*, *wzx*, *wzy* and *wbdM*. Preferred oligonucleotide molecules for use with O157 which provide for specific detection of O157 are illustrated in Table 6 and 6A.

With respect to serotypes C2 and B, suitable oligonucleotide molecules can be selected from appropriate regions described in column 3 of Tables 7 and 8.

The inventors envisage rare circumstances whereby two genetically similar gene clusters encoding serologically

different O antigens have arisen through recombination of genes or mutation so as to generate polymorphic variants. In these circumstances multiple pairs of oligonucleotides may be selected to provide hybridisation to the specific combination of genes. The invention thus provides a panel containing pairs of nucleic acid molecules for use in the method of testing of the invention, wherein the pairs of nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is specific to a particular bacterial O antigen. Preferably the particular bacterial O antigen is expressed by S. enterica. The panel of nucleic acid molecules can include pairs of nucleic acid molecules derived from sugar pathway genes where necessary.

In another aspect, the invention relates to a method for testing a food derived sample for the presence of one or more particular bacterial O antigens comprising contacting the sample with at least one pair of oligonucleotide molecules with at least one oligonucleotide molecule of the pair being capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein the gene is involved in the synthesis of the particular O antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial polysaccharide antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules. Preferably the bacteria are E. coli or S. enterica. More preferably, the E. coli are of the 0111 or 0157 serotype. More preferably the S. enterica are of the C2 or B serotype. Preferably, the method is a polymerase chain reaction method. More preferably the oligonucleotide molecules for use in the

method of the invention are labelled. Even more preferably the hybridised oligonucleotide molecules are detected by electrophoresis.

In another aspect the present invention relates to a method for testing a faecal derived sample for the presence of one or more particular bacterial O antigens comprising contacting the sample with at least one pair of oligonucleotide molecules with at least one oligonucleotide molecule of the pair being capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the particular O antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one of said genes of any bacteria expressing the particular bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules. Preferably the bacteria are E. coli or S. enterica. More preferably, the E. coli are of the 0111 or 0157 serotype. More preferably, the S. enterica are of the C2 or B serotype. Preferably, the method is a polymerase chain reaction method. More preferably the oligonucleotide molecules for use in the method of the invention are labelled. Even more preferably the hybridised oligonucleotide molecules are detected by electrophoresis.

In another aspect, the present invention relates to a method for testing a sample derived from a patient for the presence of one or more particular bacterial O antigens comprising contacting the sample with at least one pair of oligonucleotide molecules with at least one oligonucleotide molecule of the pair being capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein

said gene is involved in the synthesis of the particular O antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules. Preferably the bacteria are E. coli or S. enterica. More preferably, the E. coli are of the 0111 or 0157 serotype. More preferably, the S. enterica are of the C2 or B serotype. Preferably, the method is a polymerase chain reaction method. More preferably the oligonucleotide molecules for use in the method of the invention are labelled. Even more preferably the hybridised oligonucleotide molecules are detected by electrophoresis.

In the above described methods it will be understood that where pairs of oligonucleotides are used one of the oligonucleotide sequences may hybridise to a sequence that is not from a transferase, wzx or wzy gene. Further where both hybridise to one of these gene products they may hybridise to the same or a different one of these genes.

In addition it will be understood that where cross reactivity is an issue a combination of oligonucleotides may be chosen to detect a combination of genes to provide specificity.

The invention further relates to a diagnostic kit which can be used for the detection of bacteria which express bacterial polysaccharide antigens and the identification of the bacterial polysaccharide type of those bacteria.

Thus in a further aspect, the invention relates to a kit comprising a first vial containing a first nucleic acid molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide, including a *wzx* or wzy gene, wherein the said gene is involved in the synthesis of a bacterial polysaccharide. The kit may also provide in the same or a

separate vial a second specific nucleic acid capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide, including a *wzx* or *wzy* gene, wherein the said gene is involved in the synthesis of a bacterial polysaccharide, wherein the sequence of the second nucleic acid molecule is different from the sequence of the first nucleic acid molecule.

In a further aspect the invention relates to a kit comprising a first vial containing a first nucleic acid molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide including *wzx* or *wzy*, wherein the said gene is involved in the synthesis of a bacterial O antigen. The kit may also provide in the same or a separate vial a second specific nucleic acid capable of specifically hybridising to: (i) a gene encoding a transferase , or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide including *wzx* or *wzy*, wherein the said gene is involved in the synthesis of O antigen, wherein the sequence of the second nucleic acid molecule is different from the sequence of the first nucleic acid molecule. Preferably the first and second nucleic acid sequences are derived from E. coli or the first and second nucleic acid sequences are derived from S. enterica.

The present inventors provide full length sequence of the O157 gene cluster for the first time and recognise that from this sequence of this previously uncloned full gene cluster appropriate recombinant molecules can be generated and inserted for expression to provide expressed O157 antigens useful in applications such as vaccines.

DEFINITIONS

The phrase, "a nucleic acid molecule derived from a gene" means that the nucleic acid molecule has a

nucleotide sequence which is either identical or substantially similar to all or part of the identified gene. Thus a nucleic acid molecule derived from a gene can be a molecule which is isolated from the identified gene by physical separation from that gene, or a molecule which is artificially synthesised and has a nucleotide sequence which is either identical to or substantially similar to all or part of the identified gene. While some workers consider only the DNA strand with the same sequence as the mRNA transcribed from the gene, here either strand is intended.

Transferase genes are regions of nucleic acid which have a nucleotide sequence which encodes gene products that transfer monomeric sugar units.

Flippase or *wzx* genes are regions of nucleic acid which have a nucleotide sequence which encodes a gene product that flips oligosaccharide repeat units generally composed of three to six monomeric sugar units to the external surface of the membrane.

Polymerase or *wzy* genes are regions of nucleic acid which have a nucleotide sequence which encodes gene products that polymerise repeating oligosaccharide units generally composed of 3-6 monomeric sugar units.

The nucleotide sequences provided in this specification are described in the sequence listing as anti-sense sequences. This term is used in the same manner as it is used in Glossary of Biochemistry and Molecular Biology Revised Edition, David M. Glick, **1997** Portland Press Ltd., London on page 11 where the term is described as referring to one of the two strands of double-stranded DNA usually that which has the same sequence as the mRNA. We use it to describe this strand which has the same sequence as the mRNA.

NOMENCLATURE

Synonyms for E. coli O111 *rfb*

Current names	Our names	Bastin et al. 1991
wbdH	orf1	
gmd	orf2	
wbdI	orf3	orf3.4*
manC	orf4	rfbM*
manB	orf5	rfbK*
wbdJ	orf6	orf6.7*
wbdK	orf7	orf7.7*
wzx	orf8	orf8.9 and rfbX*
wzy	orf9	
wbdL	orf10	
wbdM	orf11	

* Nomenclature according to Bastin D.A., *et al.* 1991 "Molecular cloning and expression in Escherichia coli K-12 of the *rfb* gene cluster determining the O antigen of an E. coli O111 strain". *Mol. Microbiol.* 5:9 2223-2231.

Other Synonyms

wzy	rfc
wzx	rfbX
rmlA	rfbA
rmlB	rfbB
rmlC	rfbC
rmlD	rfbD
glf	orf6*
wbbI	orf3#, orf8* of E. coli K-12
wbbJ	orf2#, orf9* of E. coli K-12
wbbK	orf1#, orf10* of E. coli K-12
wbbL	orf5#, orf 11* of E. coli K-12

\# Nomenclature according to Yao, Z. And M. A. Valvano 1994. "Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (rfb) of Escherichia coli K-12 W3110: identification of genes the confer groups-specificty to Shigella flexineri serotypes Y and 4a". J. Bacteriol. 176: 4133-4143.

* Nomenclature according to Stevenson et al. 1994. "Structure of the O-antigen of E. coli K-12 and the sequence of its rfb gene cluster". J. Bacteriol 176: 4144-4156.

- S. enterica is a name introduced in 1987 to replace the many other names such as Salmonella typhi and Salmonella typhimurium, the old species names becoming serovar names as in S. enterica sv Typhi. However, the traditional names are still widely used.
- The O antigen genes of many species were given rfb names (rfbA etc) and the O antigen gene cluster was often referred to as the rfb cluster. There are now new names for the rfb genes as shown in the table. Both terminologies have been used herein, depending on the source of the information.

• **BRIEF DESCRIPTION OF DRAWINGS**

Figure 1 shows *Eco* R1 restriction maps of cosmid clones pPR1054, pPR1055, pPR1056, pPR1058, pPR1287 which are subclones of E. coli O111 O antigen gene cluster. The thickened line is the region common to all clones. Broken lines show segments that are non-contiguous on the chromosome. The deduced restriction map for E. coli strain M92 is shown above.

Figure 2 shows a restriction mapping analysis of E. coli O111 O antigen gene cluster within the cosmid clone pPR1058. Restriction enzymes are: (B: *Bam*H1; Bg: *Bgl*II, E: *Eco*R1; H: *Hind*III; K: *Kpn*I; P: *Pst*I; S: *Sal*I and X: *Xho*1. Plasmids pPR1230, pPR1231, and pPR1288 are deletion derivatives of pPR1058. Plasmids pPR 1237, pPR1238, pPR1239 and pPR1240 are in pUC19. Plasmids pPR1243, pPR1244, pPR1245, pPR1246 and pPR1248 are in pUC18, and pPR1292 is in pUC19. Plasmid pPR1270 is in pT7T319U. Probes 1, 2 and 3 were isolated as internal fragments of pPR1246, pPR1243 and pPR1237 respectively. Dotted lines indicate that subclone DNA extends to the left of the map into attached vector.

Figure 3 shows the structure of E. coli O111 O antigen gene cluster.

Figure 4 shows the structure of E. coli O157 O antigen gene cluster.

Figure 5 shows the structure S. enterica locus encoding the serogroup C2 O antigen gene cluster.

Figure 6 shows the structure S. enterica locus encoding the serogroup B O antigen gene cluster.

Figure 7 shows the nucleotide sequence of the E. coli O111 O antigen gene cluster. Note: (1) The first and last three bases of a gene are underlined and of italic respectively.; (2) The region which was previously sequenced by Bastin and Reeves 1995 "Sequence and anlysis of the O antigen gene (rfb) cluster of Eschericia coli o111" Gene 164: 17-23 is marked.

Figure 8 shows the nucleotide sequence of the E. coli O157 O antigen gene cluster. Note: (1) The first and last

three bases of a gene (region) are underlined and of *italic* respectively (2) The region previously sequenced by Bilge et al. 1996 "Role of the Eschericia coli O157-H7 O side chain in adherence and analysis of an rfb locus". Inf. and Immun 64:4795-4801 is marked.

Figure 9 shows the nucleotide sequence of S. enterica serogroup C2 O antigen gene cluster. Note:
(1) The numbering is as in Brown et al. 1992. "Molecular analysis of the *rfb* gene cluster of *Salmonella* serovar muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B". Mol. Microbiol. 6: 1385-1394(2) The first and last three bases of a gene are underlined and in italics respectively.(3) Only that part of the group C2 gene cluster, which differs from that of group B, was sequenced and is presented here.

Figure 10 shows the nucleotide sequence of S. enterica serogroup B O antigen gene cluster Note: (1) The numbering is as in Jiang et al. 1991. "Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2)". Mol. Microbiol. 5: 695-713. The first gene in the O antigen gene cluster is *rmlB* which starts at base 4099. (2) The first and last three bases of a gene are underlined and in italics respectively.

BEST METHOD FOR CARRYING OUT THE INVENTION

Materials and Methods-part 1

The experimental procedures for the isolation and characterisation of the E. coli O111 O antigen gene cluster (position 3,021-9,981) are according to Bastin D.A., *et al.* 1991 "Molecular cloning and expression in Escherichia coli K-12 of the *rfb* gene cluster determining the O antigen of an E. coli O111 strain". *Mol. Microbiol.* 5:9 2223-2231 and Bastin D.A. and Reeves, P.R. 1995 "Sequence and analysis of the O antigen gene(*rfb*)cluster of Escherichia coli O111". *Gene* 164: 17-23.

A. Bacterial strains and growth media

Bacteria were grown in Luria broth supplemented as required.

B. Cosmids and phage

Cosmids in the host strain x2819 were repackaged *in vivo*. Cells were grown in 250mL flasks containing 30mL of culture, with moderate shaking at 30°C to an optical density of 0.3 at 580 nm. The defective lambda prophage was induced by heating in a water bath at 45°C for 15min followed by an incubation at 37°C with vigorous shaking for 2hr. Cells were then lysed by the addition of 0.3mL chloroform and shaking for a further 10min. Cell debris were removed from 1mL of lysate by a 5min spin in a microcentrifuge, and the supernatant removed to a fresh microfuge tube. One drop of chloroform was added then shaken vigorously through the tube contents.

C. DNA preparation

Chromosomal DNA was prepared from bacteria grown overnight at 37°C in a volume of 30mL of Luria broth. After harvesting by centrifugation, cells were washed and resuspended in 10mL of 50mMTris-HCl pH 8.0. EDTA was added and the mixture incubated for 20min. Then lysozyme was added and incubation continued for a further 10min. Proteinase K, SDS, and ribonuclease were then added and the mixture incubated for up to 2hr for lysis to occur. All incubations were at 37°C. The mixture was then heated to 65°C and extracted once with 8mL of phenol at the same temperature. The mixture was extracted once with 5mL of phenol/chloroform/iso-amyl alcohol at 4°C. Residual phenol was removed by two ether extractions. DNA was precipitated with 2 vols. of ethanol at 4°C, spooled and washed in 70% ethanol, resuspended in 1-2mL of TE and dialysed. Plasmid and cosmid DNA was prepared by a modification of the Birnboim and Doly method [Birnboim, H. C. And Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA *Nucl. Acid Res.* 7:1513-1523. The volume of culture was 10mL and the lysate was extracted with phenol/chloroform/iso-amyl alcohol before precipitation with isopropanol. Plasmid

DNA to be used as vector was isolated on a continuous caesium chloride gradient following alkaline lysis of cells grown in 1L of culture.

D. Enzymes and buffers.

Restriction endonucleases and DNA T4 ligase were purchased from Boehringer Mannheim (Castle Hill, NSW, Australia)or Pharmacia LKB (Melbourne, VIC Australia). Restriction enzymes were used in the recommended commercial buffer.

E. Construction of a gene bank.

Individual aliquots of M92 chromosomal DNA (strain Stoke W, from Statens Serum Institut, 5 Artillerivej, 2300 Copenhagen S, Denmark) were partially digested with 0.2U *Sau*3A1 for 1-15mins. Aliquots giving the greatest proportion of fragments in the size range of approximately 40-50kb were selected and ligated to vector pPR691 previously digested with *Bam*H1 and *Pvu*II. Ligation mixtures were packaged *in vitro* with packaging extract. The host strain for transduction was x2819 and recombinants were selected with kanamycin.

F. Serological procedures.

Colonies were screened for the presence of the O111 antigen by immunoblotting. Colonies were grown overnight, up to 100 per plate then transferred to nitrocellulose discs and lysed with 0.5N HCl. Tween 20 was added to TBS at 0.05% final concentration for blocking, incubating and washing steps. Primary antibody was E. coli O group 111 antiserum, diluted 1:800. The secondary antibody was goat anti-rabbit IgG labelled with horseradish peroxidase diluted 1:5000. The staining substrate was 4-chloro-1-napthol. Slide agglutination was performed according to the standard procedure.

G. Recombinant DNA methods.

Restriction mapping was based on a combination of standard methods including single and double digests and sub-cloning. Deletion derivatives of entire cosmids were produced as follows: aliquots of 1.8µg of cosmid DNA were

digested in a volume of 20µl with 0.25U of restriction enzyme for 5-80min. One half of each aliquot was used to check the degree of digestion on an agarose gel. The sample which appeared to give a representative range of fragments was ligated at 4°C overnight and transformed by the $CaCl_2$ method into JM109. Selected plasmids were transformed into sϕ174 by the same method. P4657 was transformed with pPR1244 by electroporation.

H. DNA hybridisation

Probe DNA was extracted from agarose gels by electroelution and was nick-translated using [α-32P]-dCTP. Chromosomal or plasmid DNA was electrophoresed in 0.8% agarose and transferred to a nitrocellulose membrane. The hybridisation and pre-hybridisation buffers contained either 30% or 50% formamide for low and high stringency probing respectively. Incubation temperatures were 42°C and 37°C for pre-hybridisation and hybridisation respectively. Low stringency washing of filters consisted of 3 x 20min washes in 2 x SSC and 0.1% SDS. High-stringency washing consisted of 3 x 5min washes in 2 x SSC and 0.1% SDS at room temperature, a 1hr wash in 1 x SSC and 0.1% SDS at 58°C and 15min wash in 0.1 x SSC and 0.1% SDS at 58°C.

I. Nucleotide sequencing of E. coli O111 O antigen gene cluster (position 3,021-9,981)

Nucleotide sequencing was performed using an ABI 373 automated sequencer (CA, USA). The region between map positions 3.30 and 7.90 was sequenced using uni-directional exonuclease III digestion of deletion families made in PT7T3190 from clones pPR1270 and pPR1272. Gaps were filled largely by cloning of selected fragments into M13mp18 or M13mp19. The region from map positions 7.90-10.2 was sequenced from restriction fragments in M13mp18 or M13mp19. Remaining gaps in both the regions were filled by priming from synthetic oligonucleotides complementary to determined positions along the sequence,

using a single stranded DNA template in M13 or phagemid. The oligonucleotides were designed after analysing the adjacent sequence. All sequencing was performed by the chain termination method. Sequences were aligned using SAP [Staden, R., 1982 "Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing". *Nuc. Acid Res.* 10: 4731-4751; Staden, R., 1986 "The current status and portability of our sequence handling software". *Nuc. Acid Res.* 14: 217-231]. The program NIP [Staden, R. 1982 "An interactive graphics program for comparing and aligning nucleic acid and amino acid sequence". *Nuc. Acid Res.* 10: 2951-2961] was used to find open reading frames and translate them into proteins.

J. Isolation of clones carrying E. coli O111 O antigen gene cluster

The E. coli O antigen gene cluster was isolated according to the method of Bastin D.A., *et al.* [1991 "Molecular cloning and expression in Escherichia coli K-12 of the *rfb* gene cluster determining the O antigen of an E. coli O111 strain". *Mol. Microbiol.* 5(9), 2223-2231]. Cosmid gene banks of M92 chromosomal DNA were established in the *in vivo* packaging strain x2819. From the genomic bank, 3.3 x 10^3 colonies were screened with E. coli 0111 antiserum using an immuno-blotting procedure: 5 colonies (pPR1054, pPR1055, pPR1056, pPR1058 and pPR1287) were positive. The cosmids from these strains were packaged *in vivo* into lambda particles and transduced into the E. coli deletion mutant SΦ174 which lacks all O antigen genes. In this host strain, all plasmids gave positive agglutination with 0111 antiserum. An *Eco* R1 restriction map of the 5 independent cosmids showed that they have a region of approximately 11.5 kb in common (Figure 1). Cosmid pPR1058 included sufficient flanking DNA to identify several chromosomal markers linked to O antigen gene cluster and was selected for analysis of the O antigen gene cluster region.

K. Restriction mapping of cosmid pPR1058

Cosmid pPR1058 was mapped in two stages. A preliminary map was constructed first, and then the region between map positions 0.00 and 23.10 was mapped in detail, since it was shown to be sufficient for O111 antigen expression. Restriction sites for both stages are shown in Figure 2. The region common to the five cosmid clones was between map positions 1.35 and 12.95 of pPR1058.

To locate the O antigen gene cluster within pPR1058, pPR1058 cosmid was probed with DNA probes covering O antigen gene cluster flanking regions from S. enterica LT2 and E. coli K-12. Capsular polysaccharide (*cps*) genes lie upstream of O antigen gene cluster while the gluconate dehydrogenase (*gnd*) gene and the histidine (*his*) operon are downstream, the latter being further from the O antigen gene cluster. The probes used were pPR472 (3.35kb), carrying the *gnd* gene of LT2, pPR685 (5.3kb) carrying two genes of the *cps* cluster, *cpsB* and *cpsG* of LT2, and K350 (16.5kb) carrying all of the *his* operon of K-12. Probes hybridised as follows: pPR472 hybridised to 1.55kb and 3.5 kb (including 2.7 kb of vector) fragments of *Pst*1 and *Hind*III double digests of pPR1246 (a *Hind*III/*Eco*R1 subclone derived from pPR1058, Figure 2), which could be located at map positions 12.95-15.1; pPR685 hybridised to a 4.4 kb *Eco*R1 fragment of pPR1058 (including 1.3 kb of vector) located at map position 0.00-3.05; and K350 hybridised with a 32kb *Eco*R1 fragment of pPR1058 (including 4.0kb of vector), located at map position 17.30-45.90. Subclones containing the presumed *gnd* region complemented a *gnd*$^-$*edd*$^-$ strain GB23152. On gluconate bromothymol blue plates, pPR1244 and pPR1292 in this host strain gave the green colonies expected of a *gnd*$^-$*edd*$^-$ genotype. The *his*$^-$ phenotype was restored by plasmid pPR1058 in the *his* deletion strain SΦ174 on minimal medium plates, showing that the plasmid carries the entire his operon.

It is likely that the O antigen gene cluster region lies between *gnd* and *cps*, as in other E. coli and S. enterica strains, and hence between the approximate map

positions 3.05 and 12.95. To confirm this, deletion derivatives of pPR1058 were made as follows: first, pPR1058 was partially digested with *Hind*III and self ligated. Transformants were selected for kanamycin resistance and screened for expression of O111 antigen. Two colonies gave a positive reaction. *Eco*R1 digestion showed that the two colonies hosted identical plasmids, one of which was designated pPR1230, with an insert which extended from map positions 0.00 to 23.10. Second pPR1058 was digested with *Sal*1 and partially digested with *Xho*1 and the compatible ends were re-ligated. Transformants were selected with kanamycin and screened for O111 antigen expression. Plasmid DNA of 8 positively reacting clones was checked using *Eco*R1 and *Xho*1 digestion and appeared to be identical. The cosmid of one was designated pPR1231. The insert of pPR1231 contained the DNA region between map positions 0.00 and 15.10. Third, pPR1231 was partially digested with *Xho*1, self-ligated, and transformants selected on spectinomycin/ streptomycin plates. Clones were screened for kanamycin sensitivity and of 10 selected, all had the DNA region from the *Xho*1 site in the vector to the *Xho*1 site at position 4.00 deleted. These clones did not express the O111 antigen, showing that the *Xho*1 site at position 4.00 is within the O antigen gene cluster. One clone was selected and named pPR1288. Plasmids pPR1230, pPR1231, and pPR1288 are shown in Figure 2.

L. Analysis of the E. coli O111 O antigen gene cluster (position 3,021-9,981) nucleotide sequence data

Bastin and Reeves [1995 "Sequence and analysis of the O antigen gene(*rfb*)cluster of Escherichia coli O111". *Gene* 164: 17-23] partially characterised the E. coli O111 O antigen gene cluster by sequencing a fragment from map position 3,021-9,981. Figure 3 shows the gene organisation of position 3,021-9,981 of E. coli O111 O antigen gene cluster. *orf3* and *orf6* have high level amino acid identity with *wcaH* and *wcaG* (46.3% and 37.2% respectively), and are likely to be similar in function to

sugar biosynthetic pathway genes in the E. coli K-12 colanic gene cluster. *orf4* and *orf5* show high levels of amino acid homology to *manC* and *manB* genes respectively. *orf7* shows high level homology with *rfbH* which is an abequose pathway gene. *orf8* encodes a protein with 12 transmembrane segments and has similarity in secondary structure to other *wzx* genes and is likely therefore to be the O antigen flippase gene.

Materials and Methods-part 2

A. Nucleotide sequencing of 1 to 3,020 and 9,982 to 14,516 of the E. coli O111 O antigen gene cluster

The sub clones which contained novel nucleotide sequences, pPR1231 (map position 0 and 1,510), pPR1237 (map position -300 to 2,744), pPR1239 (map position 2,744 to 4,168), pPR1245 (map position 9,736 to 12,007) and pPR1246 (map position 12,007 to 15,300) (Figure 2), were characterised as follows: the distal ends of the inserts of pPR1237, pPR1239 and pPR1245 were sequenced using the M13 forward and reverse primers located in the vector. PCR walking was carried out to sequence further into each insert using primers based on the sequence data and the primers were tagged with M13 forward or reverse primer sequences for sequencing. This PCR walking procedure was repeated until the entire insert was sequenced. pPR1246 was characterised from position 12,007 to 14,516. The DNA of these sub clones was sequenced in both directions. The sequencing reactions were performed using the dideoxy termination method and thermocycling and reaction products were analysed using fluorescent dye and an ABI automated sequencer (CA, USA).

B. Analysis of the E. coli O111 O antigen gene cluster (positions 1 to 3,020 and 9,982 to 14,516 of SEQ ID NO:1) nucleotide sequence data

The gene organisation of regions of E. coli O111 O antigen gene cluster which were not characterised by Bastin and Reeves [1995 "Sequence and analysis of the O antigen gene(*rfb*)cluster of Escherichia coli O111." *Gene*

164: 17-23],(positions 1 to 3,020 and 9,982 to 14,516) is shown in Figure 3. There are two open reading frames in region 1. Four open reading frames are predicted in region 2. The position of each gene is listed in Table 5.

The deduced amino acid sequence of *orf1* (*wbdH*) shares about 64% similarity with that of the *rfp* gene of Shigella dysenteriae. Rfp and WbdH have very similar hydrophobicity plots and both have a very convincing predicted transmembrane segment in a corresponding position. *rfp* is a galactosyl transferase involved in the synthesis of LPS core, thus *wbdH* is likely to be a galactosyl transferase gene. *orf2* has 85.7% identity at amino acid level to the *gmd* gene identified in the E. coli K-12 colanic acid gene cluster and is likely to be a *gmd* gene. *orf9* encodes a protein with 10 predicted transmembrane segments and a large cytoplasmic loop. This inner membrane topology is a characteristic feature of all known O antigen polymerases thus it is likely that *orf9* encodes an O antigen polymerase gene, *wzy*. *orf10* (*wbdL*) has a deduced amino acid sequence with low homology with *Lsi2* of *Neisseria gonorrhoeae*. *Lsi2* is responsible for adding GlcNAc to galactose in the synthesis of lipooligosaccharide. Thus it is likely that *wbdL* is either a colitose or glucose transferase gene. *orf11* (*wbdM*) shares high level nucleotide and amino acid similarity with TrsE of *Yersinia enterocholitica*. TrsE is a putative sugar transferase thus it is likely that *wbdM* encodes the colitose or glucose transferase.

In summary three putative transferase genes and an O antigen polymerase gene were identified at map position 1 to 3,020 and 9,982 to 14,516 of E. coli O111 O antigen gene cluster. A search of GenBank has shown that there are no genes with significant similarity at the nucleotide sequence level for two of the three putative transferase genes or the polymerase gene. SEQ ID NO:1 and Figure 7 provide the nucleotide sequence of the O111 antigen gene cluster.

Materials and Methods-part 3

A. PCR amplification of O157 antigen gene cluster from an E. coli O157:H7 strain (Strain C664-1992, from Statens Serum Institut, 5 Artillerivej, 2300, Copenhagen S, Denmark)

E. coli O157 O antigen gene cluster was amplified by using long PCR [Cheng et al. 1994, Effective amplification of long targets from cloned inserts and human and genomic DNA" P.N.A.S. USA 91: 5695-569] with one primer (primer #412: att ggt agc tgt aag cca agg gcg gta gcg t) based on the JumpStart sequence usually found in the promoter region of O antigen gene clusters [Hobbs, *et al.* 1994 "The JumpStart sequence: a 39 bp element common to several polysaccharide gene clusted" Mol. Microbiol. 12: 855-856], and another primer #482 (cac tgc cat acc gac gac gcc gat ctg ttg ctt gg) based on the *gnd* gene usually found downstream of the O antigen gene cluster. Long PCR was carried out using the Expand Long Template PCR System from Boehringer Mannheim (Castle Hill NSW Australia), and products, 14 kb in length, from several reactions were combined and purified using the Promega Wizard PCR preps DNA purification System (Madison WI USA). The PCR product was then extracted with phenol and twice with ether, precipitated with 70% ethanol, and resuspended in 40µL of water.

B. Construction of a random DNase I bank:

Two aliquots containing about 150ng of DNA each were subjected to DNase I digestion using the Novagen DNase I Shotgun Cleavage (Madison WI USA) with a modified protocol as described. Each aliquot was diluted into 45µl of 0.05M Tris -HCl (pH7.5), 0.05mg/mL BSA and 10mM $MnCl_2$. 5µL of 1:3000 or 1:4500 dilution of DNaseI (Novagen) (Madison WI USA) in the same buffer was added into each tube respectively and 10µl of stop buffer (100mM EDTA), 30% glycerol, 0.5% Orange G, 0.075% xylene and cyanol (Novagen) (Madison WI USA) was added after incubation at 15°C for 5 min. The DNA from the two DNaseI reaction

tubes were then combined and fractionated on a 0.8% LMT agarose gel, and the gel segment with DNA of about 1kb in size (about 1.5mL agarose) was excised. DNA was extracted from agarose using Promega Wizard PCR Preps DNA Purification (Madison WI USA) and resuspended in 200 µL water, before being extracted with phenol and twice with ether, and precipitated. The DNA was then resuspended in 17.25 µL water and subjected to T4 DNA polymerase repair and single dA tailing using the Novagen Single dA Tailing Kit (Madison WI USA). The reaction product (85µl containing about 8ng DNA) was then extracted with chloroform:isoamyl alcohol (24:1) once and ligated to 3x 10^{-3} pmol pGEM-T (Promega) (Madison WI USA) in a total volume of 100µL. Ligation was carried out overnight at 4°C and the ligated DNA was precipitated and resuspended in 20µL water before being electroporated into E. coli strain JM109 and plated out on BCIG-IPTG plates to give a bank.

C. Sequencing

DNA templates from clones of the bank were prepared for sequencing using the 96-well format plasmid DNA miniprep kit from Advanced Genetic Technologies Corp (Gaithersburg MD USA) The inserts of these clones were sequenced from one or both ends using the standard M13 sequencing primer sites located in the pGEM-T vector. Sequencing was carried out on an ABI377 automated sequencer (CA USA) as described above, after carrying out the sequencing reaction on an ABI Catalyst (CA USA). Sequence gaps and areas of inadequate coverage were PCR amplified directly from O157 chromosomal DNA using primers based on the already obtained sequencing data and sequenced using the standard M13 sequencing primer sites attached to the PCR primers.

D. Analysis of the E. coli O157 O antigen gene cluster nucleotide sequence data

Sequence data were processed and analysed using the

Staden programs [Staden, R., 1982 "Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing." *Nuc. Acid Res.* 10: 4731-4751; Staden, R., 1986 "The current status and portability of our sequence handling software". *Nuc. Acid Res.* 14: 217-231; Staden, R. 1982 "An interactive graphics program for comparing and aligning nucleic acid and amino acid sequence". *Nuc. Acid Res.* 10: 2951-2961]. Figure 4 shows the structure of E. coli O157 O antigen gene cluster. Twelve open reading frames were predicted from the sequence data, and the nucleotide and amino acid sequences of all these genes were then used to search the GenBank database for indication of possible function and specificity of these genes. The position of each gene is listed in Table 6. The nucleotide sequence is presented in SEQ ID NO:2 and Figure 8.

orfs 10 and 11 showed high level identity to *manC* and *manB* and were named *manC* and *manB respectively*. *orf7* showed 89% identity (at amino acid level) to the *gmd* gene of the E. coli colanic acid capsule gene cluster (Stevenson G., K. *et al*. 1996 "Organisation of the Escherichia coli K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid".J. Bacteriol. 178:4885-4893) and was named *gmd*. *orf8* showed 79% and 69% identity (at amino acid level) respectively to *wcaG* of the E. coli colanic acid capsule gene cluster and to *wbcJ* (*orf14.8*) gene of the Yersinia enterocolitica O8 O antigen gene cluster (Zhang, L. *et al*. 1997 "Molecular and chemical characterization of the lipopolysaccharide O-antigen and its role in the virulence of Y. enterocolitica serotype O8".Mol. Microbiol. 23:63-76). Colanic acid and the Yersinia O8 O antigen both contain fucose as does the O157 O antigen. There are two enzymatic steps required for GDP-L-fucose synthesis from GDP-4-keto-6-deoxy-D-mannose, the product of the *gmd* gene product. However, it has been shown recently (Tonetti, M *et al*. 1996 Synthesis of GDP-L-fucose by the human FX protein J. Biol. Chem. 271:27274-27279) that the human FX

protein has "significant homology" with the *wcaG* gene (referred to as *Yefb* in that paper), and that the FX protein carries out both reactions to convert GDP-4-keto-6-deoxy-D-mannose to GDP-L-fucose. We believe that this makes a very strong case for *orf8* carrying out these two steps and propose to name the gene *fcl*. In support of the one enzyme carrying out both functions is the observation that there are no genes other than *manB*, *manC*, *gmd* and *fcl* with similar levels of similarity between the three bacterial gene clusters for fucose containing structures.

orf5 is very similar to *wbeE* (*rfbE*) of Vibrio cholerae 01, which is thought to be the perosamine synthetase, which converts GDP-4-keto-6-deoxy-D-mannose to GDP-perosamine (Stroeher, U.H *et al.* 1995 "A putative pathway for perosamine biosynthesis is the first function encoded within the *rfb* region of Vibrio cholerae" 01. Gene 166: 33-42). V. cholerae 01 and E. coli 0157 O antigens contain perosamine and N-acetyl-perosamine respectively. The V. cholerae 01 *manA*, *manB*, *gmd* and *wbeE* genes are the only genes of the V. cholerae 01 gene cluster with significant similarity to genes of the E. coli 0157 gene cluster and we believe that our observations both confirm the prediction made for the function of *wbe* of V. cholerae, and show that *orf5* of the 0157 gene cluster encodes GDP-perosamine synthetase. *orf5* is therefore named *per*. *orf5* plus about 100bp of the upstream region (postion 4022-5308)was previously sequenced by Bilge, S.S. *et al.* [1996 "Role of the Escherichia coli O157-H7 O side chain in adherence and analysis of an *rfb* locus".Infect. Immun. 64:4795-4801].

orf12 shows high level similarity to the conserved region of about 50 amino acids of various members of an acetyltransferase family (Lin, W., *et al.* 1994 "Sequence analysis and molecular characterisation of genes required for the biosynthesis of type 1 capsular polysaccharide in Staphylococcus aureus". J. Bateriol. 176: 7005-7016) and we believe it is the N-acetyltransferase to convert GDP-perosamine to GDP-perNAc. *orf12* has been named *wbdR*.

The genes *manB*, *manC*, *gmd*, *fcl*, *per* and *wbdR* account for all of the expected biosynthetic pathway genes of the O157 gene cluster.

The remaining biosynthetic step(s) required are for synthesis of UDP-GalNAc from UDP-Glc. It has been proposed (Zhang, L., *et al*. 1997 "Molecular and chemical characterisation of the lipopolysaccharide O-antigen and its role in the virulence of Yersinia enterocolitica serotype O8".Mol. Microbiol. 23:63-76) that in Yersinia enterocolitica UDP-GalNAc is synthesised from UDP-GlcNAc by a homologue of galactose epimerase (GalE), for which there is a *galE* like gene in the Yersinia enterocolitica O8 gene cluster. In the case of O157 there is no *galE* homologue in the gene cluster and it is not clear how UDP-GalNAc is synthesised. It is possible that the galactose epimerase encoded by the *galE* gene in the *gal* operon, can carry out conversion of UDP-GlcNAc to UDP-GalNAc in addition to conversion of UDP-Glc to UDP-Gal. There do not appear to be any gene(s) responsible for UDP-GalNAc synthesis in the O157 gene cluster.

orf4 shows similarity to many *wzx* genes and is named *wzx* and *orf2* which shows similarity of secondary structure in the predicted protein to other *wzy* genes and is for that reason named *wzy*.

The *orf1*, *orf3* and *orf6* gene products all have characteristics of transferases, and have been named *wbdN*, *wbdO* and *wbdP* respectively. The O157 O antigen has 4 sugars and 4 transferases are expected. The first transferase to act would put a sugar phosphate onto undecaprenol phosphate. The two transferases known to perform this function, WbaP (RfbP) and WecA (Rfe) transfer galactose phosphate and N-acetyl-glucosamine phosphate respectively to undecaprenol phosphate. Neither of these sugars is present in the O157 structure.

Further, none of the presumptive transferases in the O157 gene cluster has the transmembrane segments found in *WecA* and WbaP which transfer a sugar phosphate to undecaprenol phosphate and expected for any protein which

transferred a sugar to undecaprenol phosphate which is embedded within the membrane.

The *WecA* gene which transfers GlcNAc-P to undecaprenol phosphate is located in the Enterobactereal Common Antigen (ECA) gene cluster and it functions in ECA synthesis in most and perhaps all E. coli strains, and also in O antigen synthesis for those strains which have GlcNAc as the first sugar in the O unit.

It appears that WecA acts as the transferase for addition of GalNAc-1-P to undecaprenol phosphate for the Yersinia enterocolitica O8 O antigen [Zhang et al.1997 "Molecular and chemical characterisation of the lipopolysaccharide O antigen and its role in the virulence of Yersinia enterocolitica serotype O8" Mol. Microbiol. 23: 63-76.] and perhaps does so here as the O157 structure includes GalNAc. *WecA* has also been reported to add Glucose-1-P phosphate to undecaprenol phosphate in E. coli O8 and O9 strains, and an alternative possibility for transfer of the first sugar to undecaprenol phosphate is *WecA* mediated transfer of glucose, as there is a glucose residue in the O157 O antigen. In either case the requisite number of transferase genes are present if GalNAc or Glc is transferred by *WecA* and the side chain Glc is transferred by a transferase outside of the O antigen gene cluster.

orf9 shows high level similarity (44% identity at amino acid level, same length) with *wcaH* gene of the E. coli colanic acid capsule gene cluster. The function of this gene is unknown, and we give *orf9* the name *wbdQ*.

The DNA between *manB* and *wdbR* has strong sequence similarity to one of the H-repeat units of E. coli K12. Both of the inverted repeat sequences flanking this region are still recognisable, each with two of the 11 bases being changed. The H-repeat associated protein encoding gene located within this region has a 267 base deletion and mutations in various positions. It seems that the H-repeat unit has been associated with this gene cluster for a long period of time since it translocated to the gene

cluster, perhaps playing a role in assembly of the gene cluster as has been proposed in other cases.

Materials and Methods - part 4

To test our hypothesis that O antigen genes for transferases and the *wzx*, *wzy* genes were more specific than pathway genes for diagnostic PCR, we first carried out PCR using primers for all the E. coli 016 O antigen genes (Table 4). The PCR was then carried out using PCR primers for E. coli O111 transferase, *wzx* and *wzy* genes (Table 5, 5A). PCR was also carried out using PCR primers for the E. coli 0157 transferase, *wzx* and *wzy* genes (Table 6, 6A).

Chromosomal DNA from the 166 serotypes of E. coli available from Statens Serum Institut, 5 Artillerivej, 2300 Copenhagen Denmark was isolated using the Promega Genomic (Madison WI USA) isolation kit. Note that 164 of the serogroups are described by Ewing W. H.: Edwards and Ewings "Identification of the Enterobacteriacea" Elsevier, Amsterdam 1986 and that they are numbered 1-171 with numbers 31, 47, 67, 72, 93, 94 and 122 no longer valid. Of the two serogroup 19 strains we used 19ab strain F8188-41. Lior H. 1994 ["Classification of Escherichia coli In Escherichia coli in domestic animals and humans pp 31-72. Edited by C.L. Gyles CAB international] adds two more numbered 172 and 173 to give the 166 serogroups used. Pools containing 5 to 8 samples of DNA per pool were made. Pool numbers 1 to 19 (Table 1) were used in the E. coli 0111 and 0157 assay. Pool numbers 20 to 28 were also used in the 0111 assay, and pool numbers 22 to 24 contained E. coli 0111 DNA and were used as positive controls (Table 2). Pool numbers 29 to 42 were also used in the 0157 assay, and pool numbers 31 to 36 contained E. coli 0157 DNA, and were used as positive controls (Table 3). Pool numbers 2 to 20, 30, 43 and 44 were used in the E. coli 016 assay (Tables 1 to 3). Pool number 44 contained DNA of E. coli K-12 strains C600 and WG1 and was used as a positive control as between them they have all of the E.

coli K-12 O16 O antigen genes.

PCR reactions were carried out under the following conditions: denaturing 94°C/30"; annealing, temperature varies (refer to Tables 4 to 8)/30"; extension, 72°C/1'; 30 cycles. PCR reaction was carried out in an volume of 25μL for each pool. After the PCR reaction, 10μL PCR product from each pool was run on an agarose gel to check for amplified DNA.

Each E. coli and S. enterica chromosomal DNA sample was checked by gel electrophoresis for the presence of chromosomal DNA and by PCR amplification of the E. coli or S. enterica *mdh* gene using oligonucleotides based on E. coli K-12 or Salmonella enterica LT2 [Boyd *et al.* (1994) "Molecular genetic basis of allelic polymorphism in malate degydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica*" Proc. Nat. Acad. Sci. USA. 91:1280-1284.] Chromosomal DNA samples from other bacteria were only checked by gel electrophoresis of chromosomal DNA.

A. Primers based on E. coli O16 O antigen gene cluster sequence.

The O antigen gene cluster of E. coli O16 was the only typical E. coli O antigen gene cluster that had been fully sequenced prior to that of O111, and we chose it for testing our hypothesis. One pair of primers for each gene was tested against pools 2 to 20, 30 and 43 of E. coli chromosomal DNA. The primers, annealing temperatures and functional information for each gene are listed in Table 4.

For the five pathway genes, there were 17/21, 13/21, 0/21, 0/21, 0/21 positive pools for *rmlB*, *rmlD*, *rmlA*, *rmlC* and *glf* respectively (Table 4). For the *wzx*, wzy and three transferase genes there were no positives amongst the 21 pools of E. coli chromosomal DNA tested (Table 4). In each case the #44 pool gave a positive result.

B. Primers based on the E. coli 0111 O antigen gene clsuter sequence.

One to four pairs of primers for each of the transferase, *wzx* and *wzy* genes of O111 were tested against the pools 1 to 21 of E. coli chromosomal DNA (Table 5). For *wbdH*, four pairs of primers, which bind to various regions of this gene, were tested and found to be specific for O111 as there was no amplified DNA of the correct size in any of those 21 pools of E. coli chromosomal DNA tested. Three pairs of primers for *wbdM* were tested, and they are all specific although primers #985/#986 produced a band of the wrong size from one pool. Three pairs of primers for *wzx* were tested and they all were specific. Two pairs of primers were tested for *wzy*, both are specific although #980/#983 gave a band of the wrong size in all pools. One pair of primers for *wbdL* was tested and found unspecific and therefore no further test was carried out. Thus, *wzx*, *wzy* and two of the three transferase genes are highly specific to O111. Bands of the wrong size found in amplified DNA are assumed to be due to chance hybridisation of genes widely present in E. coli. The primers, annealing temperatures and positions for each gene are in (Table 5).

The O111 assay was also performed using pools including DNA from O antigen expressing Yersinia pseudotuberculosis, Shigella boydii and Salmonella enterica strains (Table 5A). None of the oligonucleotides derived from *wbdH, wzx, wzy or wbdM* gave amplified DNA of the correct size with these pools. Notably, pool number 25 includes S. enterica Adelaide which has the same O antigen as E. coli 0111: this pool did not give a positive PCR result for any primers tested indicating that these genes are highly specific for E. coli 0111.

Each of the 12 pairs binding to *wbdH, wzx, wzy and wbdM* produces a band of predicted size with the pools containing 0111 DNA (pools number 22 to 24). As pools 22 to 24 included DNA from all strains present in pool 21 plus 0111 strain DNA (Table 2), we conclude that the 12

pairs of primers all give a positive PCR test with each of three unrelated 0111 strains but not with any other strains tested. Thus these genes are highly specific for E. coli 0111.

C. Primers based on the E. coli 0157 O antigen gene cluster sequence.

Two or three primer pairs for each of the transferase, *wzx* and *wzy* genes of 0157 were tested against E. coli chromosomal DNA of pools 1 to 19, 29 and 30 (Table 6). For *wbdN*, three pairs of primers, which bind to various regions of this gene, were tested and found to be specific for 0157 as there was no amplified DNA in any of those 21 pools of E. coli chromosomal DNA tested. Three pairs of primers for *wbdO* were tested, and they are all specific although primers # 1211/#1212 produced two or three bands of the wrong size from all pools. Three pairs of primers were tested for *wbdP* and they all were specific. Two pairs of primers were tested for *wbdR* and they were all specific. For *wzy*, three pairs of primers were tested and all were specific although primer pair #1203/#1204 produced one or three bands of the wrong size in each pool. For *wzx*, two pairs of primers were tested and both were specific although primer pair #1217/#1218 produced 2 bands of wrong size in 2 pools, and 1 band of wrong size in 7 pools. Bands of the wrong size found in amplified DNA are assumed to be due to chance hybridisation of genes widely present in E. coli. The primers, annealing temperatures and function information for each gene are in Table 6.

The 0157 assay was also performed using pools 37 to 42, including DNA from O antigen expressing Yersinia pseudotuberculosis, Shigella boydii, Yersinia enterocolitica 09, Brucella abortus and Salmonella enterica strains (Table 6A). None of the oligonucleotides derived from *wbdN*, *wzy*, *wbdO*, *wzx*, *wbdP* or *wbdR* reacted specifically with these pools, except that primer pair #1203/#1204 produced two bands with Y. enterocolitica 09

and one of the bands is of the same size with that from the positive control. Primer pair #1203/#1204 binds to *wzy*. The predicted secondary structures of Wzy proteins are generally similar, although there is very low similarity at amino acid or DNA level among the sequenced *wzy* genes. Thus, it is possible that Y. enterolcolitica O9 has a *wzy* gene closely related to that of E. coli O157. It is also possible that this band is due to chance hybridization of another gene, as the other two *wzy* primer pairs (#1205/#1206 and #1207/#1208) did not produce any band with Y. enterocolitica O9. Notably, pool number 37 includes S. enterica Landau which has the same O antigen as E. coli O157, and pool 38 and 39 contain DNA of B. abortus and Y. enterocolitica O9 which cross react serologically with E. coli O157. This result indicates that these genes are highly O157 specific, although one primer pair may have cross reacted with Y. enterocolitica O9.

Each of the 16 pairs binding to *wbdN, wzx, wzy, wbdO, wbdP* and *wbdR* produces a band of predicted size with the pools containing 0157 DNA (pools number 31 to 36). As pool 29 included DNA from all strains present in pools 31 to 36 other than 0157 strain DNA (Table 3), we conclude that the 16 pairs of primers all give a positive PCR test with each of the five unrelated 0157 strains.

Thus PCR using primers based on genes *wbdN, wzy, wbdO, wzx, wbdP* and *wbdR* is highly specific for E. coli O157, giving positive results with each of six unrelated O157 strains while only one primer pair gave a band of the expected size with one of three strains with O antigens known to cross-react serologically with E. coli O157.

D. Primers based on the Salmonella enterica serotype C2 and B O antigen gene cluster sequences.

We also performed a PCR using primers for the S. enterica C2 and B serogroup transferases, *wzx, wzy* and genes (Tables 7 to 9). The nucleotide sequences of C2

and B O antigen gene clusters are listed as SEQ ID NO: 3 (Fig. 9) and SEQ ID NO:4 (Fig. 10)respectively. Chromosomal DNA from all the 46 serotypes of Salmonella enterica (Table 9) was isolated using the Promega Genomic isolation kit, 7 pools of 4 to 8 samples per pool were made. Salmonella enterica serotype B or C2 DNA was omitted from the pool for testing primers of 46 respective serotypes but added to a pool containing 6 other samples to give pool number 8 for use as a positive control.

PCR reactions were carried out under the following conditions: denaturing, 94°C/30"; annealing, temperature varies (see below)/30"; extension, 72°C/1'; 30 cycles. PCR reaction was carried out in a volume of 25µL for each pool. After the PCR reaction, 10µL PCR product from each pool was run on an agarose gel to check for amplified DNA. For pools which gave a band of correct size, PCR was repeated using individual chromosomal samples of that pool, and agarose gel was run to check for amplified DNA from each sample.

The Salmonella enterica serotype B O antigen gene cluster (of strain LT2) was the first O antigen gene cluster to be fully sequenced, and the function of each gene has been identified experimentally [Jiang, X. M., Neal, B., Santiago, F., Lee, S. J., Romana, L. K., and Reeves, P. R. (1991) "Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2)." Mol. Microbiol. **5**(3), 695-713; Liu, D., Cole, R., and Reeves, P. R. (1996). "An O antigen processing function for Wzx(RfbX): a promising candidate for O-unit flippase" J. Bacteriol., **178**(7),2102-2107; Liu, D., Haase, A. M., Lindqvist, L., Lindberg, A. A., and Reeves, P. R. (1993). "Glycosyl transferases of O-antigen biosynthesis in *S. enterica* : identification and characterisation of transferase genes of groups B, C2 and E1." J. Bacteriol., **175**, 3408-3413; Liu, D., Lindquist, L., and Reeves P. R. (1995). "Transferases of O-antigen biosynthesis in *Salmonella enterica*: dideoxhexosyl

transferases of groups B and C2 and acetyltransferase of group C2." J. Bacteriol., **177**, 4084-4088; Romana, L. K., Santiago, F. S., and Reeves, P. R. (1991). "High level expression and purification dThymidine-diphospho-D-glucose 4,6 dehydratase (*rfbB*) from *Salmonella* serovar *typhimurium* LT2." BBRC, **174**, 846-852]. One pair of primers for each of the pathway genes and *wbaP* was tested against the pools of Salmonella enterica DNA, two to three pairs of primers for each of the other transferases and *wzx* genes were also tested. See Table 8 for a list of primers and functional information of each gene, as well as the annealing temperature of the PCR reaction for each pair of primers.

For pathway genes of group B strain LT2, there are 19/45, 14/45, 15/45, 12/45, 6/45, 6/45, 6/45, 6/45, 1/45, 9/45, 8/45 positives for *rmlB*, *rmlD*, *rmlA*, *rmlC*, *ddhD*, *ddhA*, *ddhB*, *ddhC*, *abe*, *manC*, and *manB* repsectively (Table 9).

For the LT2 *wzx* gene we used three primer pairs each of which gave 1/45 positive. For the 4 transferase genes we used a total of 9 primer pairs. 2 primer pairs for *wbaV* gave 2/90 positives. For 3 primer pairs of *wbaN*, 11/135 gave a positive result. For the *wbaP* primer pair 10/45 gave a positive result (Table 9).

The experimental data show that oligonucleotides derived from the *wzx* and *wbaV* group B O antigen genes are specific for group B O antigen amongst all 45 Salmonella enterica O antigen groups except O group 67. The oligonucleotides derived from Salmonella enterica B group *wbaN* and *wbaU* genes detected B group O antigen and also produced positive results with groups A, D1 and D3. *WbaU* encodes a transferase for a Mannose α(1-4) Mannose linkage and is expressed in groups A, B and D1 while *wbaN*, which encodes a transferase for Rhamnose α(1-3) Galactose linkage is present in groups A, B, D1, D2, D3 and E1. This accounts for the positive results with the group B *wbaU* and *wbaN* genes. The *wbaN* gene of groups E and D2 has considerable sequence differences from that of groups A,

B, D1 and D3 and this accounts for the positive results only with groups B, D1 and D3.

The Salmonella enterica B primers derived from *wzx* and transferase genes produced a positive result with Salmonella enterica 067. We find that Salmonella enterica O67 has all the genes of the group B O antigen cluster. There are several possible explanations for this finding including the possibility that the gene cluster is not functional due to mutation and the group 067 antigenicity is due to another antigen, or the O antigen is modified after synthesis such that its antigenicity is changed. Salmonella enterica 067 would therefore be scored as Salmonella enterica group B in the PCR diagnostic assay. However, this is of little importance because Salmonella enterica 067 is a rare O antigen and only one (serovar Crossness) of the 2324 known serovars has the 067 serotype [Popoff M.Y. *et al* (1992) "Antigenic formulas of the Salmonella enterica serovars" 6th revision WHO Collaborating Centre for Reference and Research on Salmonella enterica, Institut Pasteur Paris France], and serovar Crossness had only been isolated once [M. Popoff, personal communication].

The Salmonella enterica B primers derived from *wbaP* reacted with group A, C2, D1, D2, D3, E1, 54, 55, 67 and E4 O antigen groups. *WbaP* encodes the galactosyl transferase which initiates O unit synthesis by transfer of Galactose phosphate to the lipid carrier Undecaprenol phosphate. This reaction is common to the synthesis of several O antigens. As such *wbaP* is distinguished from other transferases of the invention as it does not make a linkage within an O antigen.

We also tested 20 primer pairs for the *wzx*, *wzy* and 5 transferase genes of serotype C2 and found no positives in all the 7 pools (Table 7).

Groups A, B, D1, D2, D3, C2 and E1 share many genes in common. Some of these genes occur with more than one sequence in which case each specific sequence can be named after one of the serogroups in which it occurs. The

distribution of these sequence specificities is shown in Table 10. The inventors have aligned the nucleotide sequences of Salmonella enterica *wzy*, *wzx* genes and transferase genes so as to determine specific combinations of nucleic acid molecules which can be employed to specifically detect and identify the Salmonella enterica groups A, B, D1, D2, D3, C2 and E1 (Table 10). The results show that many of the O antigen groups can be detected and identified using a single specific nucleic acid molecule although other groups in particular D2 and E1, and A and D1 require a panel of nucleic acid molecules derived from a combination of genes.

It will be understood that in carrying out the methods of the invention with respect to the testing of particular sample types including samples from food, patients and faeces the samples are prepared by routine techniques routinely used in the preparation of such samples for DNA based testing.

TABLE 1

Pool No.	Strains of which chromosonal DNA included in the pool	Source*
1	*E. coli* type strains for O serotypes 1, 2, 3, 4, 10, 16, 18 and 39	IMVS[a]
2	*E. coli* type strains for O serotypes 40, 41, 48, 49, 71, 73, 88 and 100	IMVS
3	*E. coli* type strains for O serotypes 102, 109, 119, 120, 121, 125, 126 and 137	IMVS
4	*E. coli* type strains for O serotypes 138, 139, 149, 7, 5, 6, 11 and 12	IMVS
5	*E. coli* type strains for O serotypes 13, 14, 15, 17, 19ab, 20, 21 and 22	IMVS
6	*E. coli* type strains for O serotypes 23, 24, 25, 26, 27, 28, 29 and 30	IMVS
7	*E. coli* type strains for O serotypes 32, 33, 34, 35, 36, 37, 38 and 42	IMVS
8	*E. coli* type strains for O serotypes 43, 44, 45, 46, 50, 51, 52 and 53	IMVS
9	*E. coli* type strains for O serotypes 54, 55, 56, 57, 58, 59, 60 and 61	IMVS
10	*E. coli* type strains for O serotypes 62, 63, 64, 65, 66, 68, 69 and 70	IMVS
11	*E. coli* type strains for O serotypes 74, 75, 76, 77, 78, 79, 80 and 81	IMVS
12	*E. coli* type strains for O serotypes 82, 83, 84, 85, 86, 87, 89 and 90	IMVS
13	*E. coli* type strains for O serotypes 91, 92, 95, 96, 97, 98, 99 and 101	IMVS
14	*E. coli* type strains for O serotypes 103, 104, 105, 106, 107, 108 and 110	IMVS
15	*E. coli* type strains for O serotypes 112, 162, 113, 114, 115, 116, 117 and 118	IMVS
16	*E. coli* type strains for O serotypes 123, 165, 166, 167, 168, 169, 170 and 171	See b
17	*E. coli* type strains for O serotypes 172, 173, 127, 128, 129, 130, 131 and 132	See c
18	*E. coli* type strains for O serotypes 133, 134, 135, 136, 140, 141, 142 and 143	IMVS
19	*E. coli* type strains for O serotypes 144, 145, 146, 147, 148, 150, 151 and 152	IMVS

*

a. Institute of Medical and Veterinary Science, Adelaide, Australia

b. 123 from IMVS; the rest from Statens Serum Institut, Copenhagen, Denmark

c. 172 and 173 from Statens Serum Institut, Copenhagen, Denmark, the rest from IMVS

TABLE 2

Pool No.	Strains of which chromosonal DNA included in the pool	Source*
20	*E. coli* type strains for O serotypes 153, 154, 155, 156, 157, 158 , 159 and 160	IMVS
21	*E. coli* type strains for O serotypes 161, 163, 164, 8, 9 and 124	IMVS
22	As pool #21, plus *E. coli* 0111 type strain Stoke W.	IMVS
23	As pool #21, plus *E. coli* 0111:H2 strain C1250-1991	See d
24	As pool #21, plus *E. coli* 0111:H12 strain C156-1989	See e
25	As pool #21, plus *S. enterica* serovar Adelaide	See f
26	*Y. pseudotuberculosis* strains of O groups IA, IIA, IIB, IIC, III, IVA, IVB, VA, VB, VI and VII	See g
27	*S. boydii* strains of serogroups 1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14 and 15	See h
28	*S. enterica* strains of serovars (each representing a different O group) Typhi, Montevideo, Ferruch, Jangwani, Raus, Hvittingfoss, Waycross, Dan, Dugbe, Basel, 65,:i:e,n,z,15 and 52:d:e,n,x,z15	IMVS

*

d. C1250-1991 from Statens Serum Institut, Copenhagen, Denmark
e. C156-1989 from Statens Serum Institut, Copenhagen, Denmark
f. *S. enterica* serovar Adelaide from IMVS
g. Dr S Aleksic of Institute of Hygiene, Germany
h. Dr J Lefebvre of Bacterial Identification Section, Laboratoroie de Santè Publique du Quèbec, Canada

TABLE 3

Pool No.	Strains of which chromosonal DNA included in the pool	Source*
29	*E. coli* type strains for O serotypes 153, 154, 155, 156, 158, 159 and 160	IMVS
30	*E. coli* type strains for O serotypes 161, 163, 164, 8, 9, 111 and 124	IMVS
31	As pool #29, plus *E. coli* O157 type strain A2 (O157:H19)	IMVS
32	As pool #29, plus *E. coli* O157:H16 strain C475-89	See d
33	As pool #29, plus *E. coli* O157:H45 strain C727-89	See d
34	As pool #29, plus *E. coli* O157:H2 strain C252-94	See d
35	As pool #29, plus *E. coli* O157:H39 strain C258-94	See d
36	As pool #29, plus *E. coli* O157:H26	See e
37	As pool #29, plus *S. enterica* serovar Landau	See f
38	As pool #29, plus *Brucella abortus*	See g
39	As pool #29, plus *Y. enterocolitica* O9	See h
40	*Y. pseudotuberculosis* strains of O groups IA, IIA, IIB, IIC, III, IVA, IVB, VA, VB, VI and VII	See i
41	*S. boydii* strains of serogroups 1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14 and 15	See j
42	*S. enterica* strains of serovars (each representing a different O group) Typhi, Montevideo, Ferruch, Jangwani, Raus, Hvittingfoss, Waycross, Dan, Dugbe, Basel, 65:i:e,n,z15 and 52:d:e,n,x,z15	IMVS
43	*E. coli* type strains for O serotypes 1,2,3,4,10,18 and 29	IMVS
44	As pool #43, plus *E. coli* K-12 strains C600 and WG1	IVMS See k

*

d. O157 strains from Statens Serum Institut, Copenhagen, Denmark
e. O157:H26 from Dr R Brown of Royal Children's Hospital, Melbourne, Victoria
f. *S. enterica* serovar Landau from Dr M Poppoff of Institut Pasteur, Paris, France
g. *B. Abortus* from the culture collection of The University of Sydney, Sydney, Australia
h. *Y. enterocolitica* O9 from Dr. K. Bettelheim of Victorian Infectious Diseases Reference Laboratory Victoria, Australia.
i. Dr S Aleksic of Institute of Hygiene, Germany
J. Dr J Lefebvre of Bacterial Identification Section, Laboratoroie de Santè Publique du Quèbec, Canada
k. Strains C600 and WG1 from Dr. B.J. Backmann of Department of Biology, Yale University, USA.

TABLE 4 PCR assay result using primers based on the E. coli serotype O16 (strain K-12) O antigen gene cluster sequence

Gene	Function	Base positions of the gene	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (out of 21) giving band of correct size	Annealing temperature of the PCR
*rmlB**	TDP-rhamnose pathway	90-1175	#1064(91-109)	#1065(1175-1157)	1085bp	17	60°C
*rmlD**	TDP-rhamnose pathway	1175-2074	#1066(1175-1193)	#1067 (2075-2058)	901bp	13	60°C
*rmlA**	TDP-rhamnose pathway	2132-3013	#1068(2131-2148)	#1069(3013-2995)	883bp	0	60°C
*rmlC**	TDP-rhamnose pathway	3013-3570	#1070(3012-3029)	#1071(3570-3551)	559bp	0	60°C
*gtf**	Galactofuranose pathway	4822-5925	#1074(4822-4840)	#1075(5925-5908)	1104bp	0	55°C
*wzx**	Flippase	3567-4814	#1072(3567-3586)	#1073(4814-4797)	1248bp	0	55°C
*wzy**	O polymerase	5925-7091	#1076(5925-5944)	#1077(7091-7074)	1167bp	0	60°C
*wbbI**	Galactofuranosyl transferase	7094-8086	#1078 (7094-7111)	#1079(8086-8069)	993bp	0	50°C
*wbbJ**	Acetyltransferase	8067-8654	#1080(8067-8084)	#1081(8654-8632)	588bp	0	60°C
*wbbK***	Glucosyl transferase	5770-6888	#1082(5770-5787)	#1083(6888-6871)	1119bp	0	55°C
*wbbL****	Rhamanosyltransferase	679-1437	#1084(679-697)	#1085(1473-1456)	795bp	0****	55°C

*, **, *** Base positions based on GenBank entry U09876, U03041 and L19537 respectively

**** 19 pools giving a band of wrong size

TABLE 5 PCR assay data using 0111 primers

Gene	Base positions of the gene according to SEQ ID NO: 1	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (out of 21) giving band of correct size	Annealing temperature of the PCR
wbdH	739-1932	#866 (739-757)	#867(1941-1924)	1203bp	0	60°C
		#976(925-942)	#978(1731-1714)	807bp	0	60°C
		#976(925-942)	#979(1347-1330)	423bp	0	60°C
		#977(1165-1182)	#978(1731-1714)	567bp	0	60°C
wzx	8646-9911	#969(8646-8663)	#970(9908-9891)	1263bp	0	50°C
		#1060(8906-8923)	#1062(9468-9451)	563bp	0	60°C
		#1061(9150-9167)	#1063 (9754-9737)	605bp	0	50°C
wzy	9901-10953	#900(9976-9996)	#901(10827-10807)	852bp	0	60°C
		#980(10113-10130)	#983(10484-10467)	372bp	0*	61°C
wbdL	10931-11824	#870(10931-10949)	#871(11824-11796)	894bp	7	60°C
wbdM	11821-12945	#868(11821-11844)	#869(12945-12924)	1125bp	0	60°C
		#984(12042-12059)	#987(12447-12430)	406bp	0	60°C
		#985(12258-12275)	#986(12698-12681)	441bp	0**	65°C

* Giving a band of wrong size in all pools

** One pool giving a band of wrong size

TABLE 5A PCR specificity test data using 0111 primers

Gene	Base positions of the gene according to SEQ ID NO: 1	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (pools no. 25-28) giving band of correct size	Annealing temperature of the PCR
wbdH	739-1932	#866 (739-757)	#867(1941-1924)	1203bp	0*	60°C
		#976(925-942)	#978(1731-1714)	807bp	0	60°C
		#976(925-942)	#979(1347-1330)	423bp	0	60°C
		#977(1165-1182)	#978(1731-1714)	567bp	0	60°C
wzx	8646-9911	#969(8646-8663)	#970(9908-9891)	1263bp	0	55°C
		#1060(8906-8923)	#1062(9468-9451)	563bp	0	60°C
		#1061(9150-9167)	#1063 (9754-9737)	605bp	0*	50°C
wzy	9901-10953	#900(9976-9996)	#901(10827-10807)	852bp	0	60°C
		#980(10113-10130)	#983(10484-10467)	372bp	0**	60°C
wbdL	10931-11824	#870(10931-10949)	#871(11824-11796)	894bp	0	60°C
wbdM	11821-12945	#868(11821-11844)	#869(12945-12924)	1125bp	0	60°C
		#984(12042-12059)	#987(12447-12430)	406bp	0	60°C
		#985(12258-12275)	#986(12698-12681)	441bp	0*	65°C

* 1 pool giving a band of wrong size

** 2 pools giving 3 bands of wrong sizes, 1 pool giving 2 bands of wrong sizes

TABLE 6 PCR results using primers based on the *E. coli* O157 sequence

Gene	Function	Base position of the gene according to SEQ ID NO: 2	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (out of 21) giving band of correct size	Annealing temperature of the PCR
wbdN	Sugar transferase	79-861	#1197(79-96)	#1198 (861-844)	783	0	55°C
			#1199(184-201)	#1200(531-514)	348	0	55°C
			#1201(310-327)	#1202(768-751)	459	0	55°C
wzy	O antigen	858-2042	#1203(858-875)	#1204(2042-2025)	1185	0*	50°C
			#1205(1053-1070)	#1206(1619-1602)	567	0	63°C
			#1207(1278-1295)	#1208(1913-1896)	636	0	60°C
wbdO	Sugar transferase	2011-2757	#1209(2011-2028)	#1210(2757-2740)	747	0	50°C
			#1211(2110-2127)	#1212(2493-2476)	384	0**	62°C
			#1213(2305-2322)	#1214(2682-2665)	378	0	60°C
wzx	O antigen flippase	2744-4135	#1215(2744-2761)	#1216(4135-4118)	1392	0	50°C
			#1217(2942-2959)	#1218(3628-3611)	687	O***	63°C
wbdP	Sugar transferase	5257-6471	#1221(5257-5274)	#1222(6471-6454)	1215	0	55°C
			#1223(5440-5457)	#1224(5973-5956)	534	0	55°C
			#1225(5707-5724)	#1226(6231-6214)	525	0	55°C
wbdR	N-acetyl transferase	13156-13821	#1229(13261-13278)	#1230(13629-13612)	369	0	55°C
			#1231(13384-13401)	#1232(13731-13714)	348	0	60°C

* 3 bands of wrong size in one pool, 1 band of wrong size in all other pools

** 3 bands of wrong sizes in 9 pools, 2 bands of wrong size in all other pools

*** 2 bands of wrong sizes in 2 pools, 1 band of wrong size in 7 pools

TABLE 6A PCR results using primers based on the *E. coli* O157 sequence

Gene	Function	Base position of the gene according to SEQ ID NO: 2	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (pools no. 37-42) giving band of correct size	Annealing temperature of the PCR
wbdN	Sugar transferase	79-861	#1197(79-96)	#1198 (861-844)	783	0*	55°C
			#1199(184-201)	#1200(531-514)	348	0*	55°C
			#1201(310-327)	#1202(768-751)	459	0	61°C
wzy	O antigen polymerase	858-2042	#1203(858-875)	#1204(2042-2025)	1185	1**	50°C
			#1205(1053-1070)	#1206(1619-1602)	567	0***	60°C
			#1207(1278-1295)	#1208(1913-1896)	636	0	60°C
wbdO	Sugar transferase	2011-2757	#1209(2011-2028)	#1210(2757-2740)	747	0	50°C
			#1211(2110-2127)	#1212(2493-2476)	384	0****	61°C
			#1213(2305-2322)	#1214(2682-2665)	378	0	60°C
wzx	O antigen flippase	2744-4135	#1215(2744-2761)	#1216(4135-4118)	1392	0	50°C
			#1217(2942-2959)	#1218(3628-3611)	687	0	63°C
wbdP	Sugar transferase	5257-6471	#1221(5257-5274)	#1222(6471-6454)	1215	0	55°C
			#1223(5440-5457)	#1224(5973-5956)	534	0*	60°C
			#1225(5707-5724)	#1226(6231-6214)	525	0	55°C
wbdR	N-acetyl transferase	13156-13821	#1229(13261-13278)	#1230(13629-	369	0	50°C
			#1231(13384-13401)	#1232(13731-	348	0	60°C

* 1 band of wrong size in one pool

** pool #39 giving two bands, one band of correct size, the other band of wrong size in another pool.

*** 2 bands of wrong sizes in one pool

**** 3 bands of wrong sizes in 2 pools, 2 bands of wrong sizes in 2 other pools

TABLE 7

PCR assay data using primers based on the *Salmonella enterica* serotype C2 (strain M67) O antigen gene cluster sequence

Gene	Function	Base positions of the gene according to SEQ ID NO: 3	Forward primer (base position)	Reverse primer (base position)	Length of the PCR fragment	Number of pools (out of 7) giving band of correct size	Annealing temperature of the PCR
wzx	Flippase	1019-2359	#1144(1019-1036)	#1145(1414-1397)	396bp	0	55°C
			#1146(1708-1725)	#1147(2170-2153)	463bp	0	55°C
			#1148(1938-1955)	#1149(2356-2339)	419bp	0	55°C
wbaR	Abequosyl transferase	2352-3314	#1150(2332-2369)	#1151(2759-2742)	408bp	0	55°C
			#1152(2601-2618)	#1153(3047-3030)	447bp	0	55°C
			#1154(2910-2927)	#1155(3311-3294)	402bp	0	55°C
wbaL	Acetyl transferase	3361-3875	#1156(3361-3378)	#1157(3759-3742)	399bp	0	55°C
			#1158(3578-3595)	#1159(3972-3955)	395bp	0	50°C
wbaQ	Rhamnosyl	3977-5020	#1160(3977-3994)	#1161(4378-4361)	402bp	0	55°C
			#1162(4167-4184)	#1163(4774-4757)	608bp	0	55°C
			#1164(4603-4620)	#1165(5017-5000)	415bp	0*	60°C
wzy	O polymerase	5114-6313	#1166(5114-5131)	#1167(5515-5498)	402bp	0**	55°C
			#1168(5664-5681)	#1169(6112-6095)	449bp	0	55°C
			#1170(5907-5924)	#1171(6310-6293)	404bp	0	55°C
wbaW	Mannosyl transferase	6313-7323	#1172(6313-6330)	#1173(6805-6788)	493bp	0	50°C
			#1174(6697-6714)	#1175(7068-7051)	372bp	0	55°C
			#1176(6905-6922)	#1177(7320-7303)	416bp	0	55°C
wbaZ	Mannosyl transferase	7310-8467	#1178(7310-7327)	#1179(7775-7758)	466bp	0	50°C
			#1180(7530-7547)	#1181(7907-7890)	378bp	0	55°C
			#1182(8007-8024)	#1183(8464-8447)	458bp	0	55°C

* Positive pool gives another band, which is also present in another pool. All other pools gave bands of wrong size.

** Band of wrong size in 6 other pools.

TABLE 8

PCR primers based on the *Salmonella enterica* serotype B (strain LT2) O antigen gene cluster sequence

Gene	Function	Base position of the gene according to SEQ ID NO: 4	Forward primer (base position)	Reverse primer (base position)	Length of the PCR fragment	Annealing temperature of the PCR
rmlB	TDP-rhamnose pathway	4099-5184	#1094 (4100-4117)	#1095(4499-4482)	400bp	55°C
rmlD	TDP-rhamnose pathway	5184-6083	#1092(5186-5203)	#1093(5543-5526)	358bp	50°C
rmlA	TDP-rhamnose pathway	6131-7009	#1090(6531-6548)	#1091(6837-6820)	308bp	55°C
rmlC	TDP-rhamnose pathway	7010-7561	#1088(7013-7030)	#1089(7372-7355)	360bp	55°C
ddhD	CDP-abequose pathway	7567-8559	#1112(7567-7584)	#1113(7970-7953)	404bp	55°C
ddhA	CDP-adequose pathway	8556-9329	#1114(8556-8573)	#1115(8975-8958)	420bp	60°C
ddhB	CDP-adequose pathway	9334-10413	#1116(9334-9351)	#1117(9816-9799)	483bp	45°C
ddhC	CDP-adequose pathway	10440-11753	#1118(10440-10457)	#1119(10871-10854)	432bp	60°C
abe	CDP-adequose pathway	11781-12680	#1100(12008-12025)	#1101(12388-12371)	381bp	55°C
wzx	Flippase	12762-14054	#1120(12762-12779)	#1121(13150-13133)	389bp	55°C
			#1122(12993-13010)	#1123(13417-13400)	425bp	55°C
			#1124(13635-13652)	#1125(14051-14034)	417bp	55°C
wbaV	Abequosyl transferase	14059-15060	#1126(14059-14076)	#1127(14421-14404)	363bp	45°C
			#1128(14688-14705)	#1129(15057-15040)	370bp	45°C
wbaU	Mannosyl transferase	15379-16440	#1130(15379-15396)	#1131(15768-15751)	390bp	60°C
			#1132(15850-15867)	#1133(16262-16245)	413bp	50°C
			#1134(16027-16044)	#1135(16437-16420)	411bp	60°C
wbaN	Rhamnosyl transferase	16441-17385	#1136(16441-16458)	#1137(16851-16834)	411bp	45°C
			#1138(16630-16647)	#1139(17087-17070)	458bp	55°C
			#1140(16978-16995)	#1141(17382-17365)	405bp	50°C
manC	GDP-mannose pathway	17386-18825	#1098(17457-17474)	#1099(18143-18126)	687bp	60°C
mabB	GDP-mannose pathway	18812-20245	#1096(18991-19008)	#1097(19345-19328)	355bp	55°C
wbaP	Galactosyl transferase	20317-21747	#1142(20389-20406)	#1143(20709-20692)	321bp	55°C

TABLE 9 PCR results using LT2 primers*

Strain name	O group	1094 - 1095	1092 - 1093	1090 - 1091	1088 - 1089	1112 - 1113	1114 - 1115	1116 - 1117	1118 - 1119	1100 - 1101	1120 - 1121	1122 - 1123	1124 - 1125	1126 - 1127	1128 - 1129	1130 - 1131	1132 - 1133	1134 - 1135	1136 - 1137	1138 - 1139	1140 - 1141	1098 - 1099	1096 - 1097	1142 - 1143
		rmlB	rmlD	rmlA	rmlC	ddhD	ddhA	ddhB	ddhC	abe	wzx	wzx	wzx	wbaV	wbaV	wbaU	wbaU	wbaU	wbaN	wbaN	wbaN	manC	manB	wbap
M8	A	y	y	y	y	y	y	y	y							y	y	y	y	y	y	y	y	y
P5003	B	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
M40	C1																					y	y	y
M47	C2	y	y	y	y	y	y	y	y							y	y	y	y	y	y	y	y	y
M10	D1	y	y	y	y	y	y	y	y													y	y	y
M388	D2	y		y	y	y	y	y	y															
M840	D3	y	y	y	y	y	y	y	y							y	y			y	y	y	y	y
M31	E1	y	y	y	y																			
M324	F	y	y	y	y																			
M329	G																							
M282	H																							
M284	I																							
M384	J																							
M285	K																							
M7	L																							
M289	M	y	y	y																				
M270	N																							
M99	O																							
M360	P																							
M273	Q	y																						
M261	R																							
M382	S																							
M387	T	y	y	y	y																			
M295	U																							
M289	V																							
M396	W																							
M278	X																							
M390	Y																							
M93	Z																							
M291	51																							
M309	53																							
M303	53	y	y	y																	y	y	y	
M295	54	y	y	y	y																			
M304	56	y																						
M293	57	y	y	y																				
M305	58																							
M385	59	y	y	y	y																			
M306	60	y																						
M328	61																							
M330	65																							
M322	66																							y
M1408	55																							
M1409	62																							
M1410	63	y																						
M1411	67	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
M74	E4	y	y	y	y																			

* y indicates a positive PCR result. Blank indicates a negative result.

TABLE 10 Gene specificities in *Salmonella enterica* serogroups

Serogroup	Genes										
	wzy	wzx	wbaP	wbaU	wbaN	wbaV	wbaO	wbaW	wbaZ	wbaQ	wbaR
A	B	D	B	B	B	D	-	-	-	-	-
B	B	B	B	B	B	B	-	-	-	-	-
D1	B	D	B	B	B	D	-	-	-	-	-
D2	E1	D	B	-	E1	D	E1	-	-	-	-
D3	D3	D	B	B	B	D	-	-	-	-	-
C2	C2	C2	B	-	-	-	-	C2	C2	C2	C2
E1	E1	E1	B	-	E1	-	E1				

- means 'not present'

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Reeves, Peter R
Wang, Lei

(ii) TITLE OF INVENTION: Nucleic Acid Molecules Specific For Bacterial Antigens And Uses Thereof

(iii) NUMBER OF SEQUENCES: 4

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Thomas Gumley
(B) STREET: 168 Walker Street
(C) CITY: North Sydney
(D) STATE: New South Wales
(E) COUNTRY: Australia
(F) ZIP: 2068

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Gumley, Thomas P

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 99575944
(B) TELEFAX: 99576288

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14516 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(v) ORIGINAL SOURCE:
(A) ORGANISM: Escherichia coli

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GATCTGATGG CCGTAGGGCG CTACGTGCTT TCTGCTGATA TCTGGGCTGA GTTGGAAAAA 60

ACTGCTCCAG GTGCCTGGGG ACGTATTCAA CTGACTGATG CTATTGCAGA GTTGGCTAAA 120

AAACAGTCTG TTGATGCCAT GCTGATGACC GGCGACAGCT ACGACTGCGG TAAGAAGATG 180

GGCTATATGC AGGCATTCGT TAAGTATGGG CTGCGCAACC TTAAAGAAGG GGCGAAGTTC 240
CGTAAGAGCA TCAAGAAGCT ACTGAGTGAG TAGAGATTTA CACGTCTTTG TGACGATAAG 300
CCAGAAAAAA TAGCGGCAGT TAACATCCAG GCTTCTATGC TTTAAGCAAT GGAATGTTAC 360
TGCCGTTTTT TATGAAAAAT GACCAATAAT AACAAGTTAA CCTACCAAGT TTAATCTGCT 420
TTTTGTTGGA TTTTTTCTTG TTTCTGGTCG CATTTGGTAA GACAATTAGC GTGAGTTTTA 480
GAGAGTTTTG CGGGATCTCG CGGAACTGCT CACATCTTTG GCATTTAGTT AGTGCACTGG 540
TAGCTGTTAA GCCAGGGGCG GTAGCTTGCC TAATTAATTT TTAACGTATA CATTTATTCT 600
TGCCGCTTAT AGCAAATAAA GTCAATCGGA TTAAACTTCT TTTCCATTAG GTAAAAGAGT 660
GTTTGTAGTC GCTCAGGGAA ATTGGTTTTG GTAGTAGTAC TTTTCAAATT ATCCATTTTC 720
CGATTTAGAT GGCAGTTGAT GTTACTATGC TGCATACATA TCAATGTATA TTATTTACTT 780
TTAGAATGTG ATATGAAAAA AATAGTGATC ATAGGCAATG TAGCGTCAAT GATGTTAAGG 840
TTCAGGAAAG AATTAATCAT GAATTTAGTG AGGCAAGGTG ATAATGTATA TTGTCTAGCA 900
AATGATTTTT CCACTGAAGA TCTTAAAGTA CTTTCGTCAT GGGGCGTTAA GGGGGTTAAA 960
TTCTCTCTTA ACTCAAAGGG TATTAATCGT TTTAAGGATA TAATTGCTGT TTATGAACTA 1020
AAAAAAATTC TTAAGGATAT TTCCCCAGAT ATTGTATTTT CATATTTTGT AAAGCCAGTA 1080
ATATTTGGAA CTATTGCTTC AAAGTTGTCA AAAGTGCCAA GGATTGTTGG AATGATTGAA 1140
GGTCTAGGTA ATGCCTTCAC TTATTATAAG GGAAAGCAGA CCACAAAAAC TAAAATGATA 1200
AAGTGGATAC AAATTCTTTT ATATAAGTTA GCATTACCGA TGCTTGATGA TTTGATTCTA 1260
TTAAATCATG ATGATAAAAA AGATTTAATC GATCAGTATA ATATTAAAGC TAAGGTAACA 1320
GTGTTAGGTG GGATTGGATT GGATCTTAAT GAGTTTTCAT ATAAAGAGCC ACCGAAAGAG 1380
AAAATTACCT TTATTTTTAT AGCAAGGTTA TTAAGAGAGA AAGGGATATT TGAGTTTATT 1440
GAAGCCGCAA AGTTCGTTAA GACAACTTAT CCAAGTTCTG AATTTGTAAT TTTAGGAGGT 1500
TTTGAGAGTA ATAATCCTTT CTCATTACAA AAAAATGAAA TTGAATCGCT AAGAAAAGAA 1560
CATGATCTTA TTTATCCTGG TCATGTGGAA AATGTTCAAG ATTGGTTAGA GAAAAGTTCT 1620
GTTTTTGTTT TACCTACATC ATATCGAGAA GGCGTACCAA GGGTGATCCA AGAAGCTATG 1680
GCTATTGGTA GACCTGTAAT AACAACTAAT GTACCTGGGT GTAGGGATAT AATAAATGAT 1740
GGGGTCAATG GCTTTTTGAT ACCTCCATTT GAAATTAATT TACTGGCAGA AAAAATGAAA 1800
TATTTTATTG AGAATAAAGA TAAAGTACTC GAAATGGGGC TTGCTGGAAG GAAGTTTGCA 1860
GAAAAAAACT TTGATGCTTT TGAAAAAAAT AATAGACTAG CATCAATAAT AAAATCAAAT 1920
AATGATTTTT GACTTGAGCA GAAATTATTT ATATTTCAAT CTGAAAAATA AAGGCTGTTA 1980
TTATGAATAA AGTGGCATTA ATTACTGGTA TCACTGGGCA AGATGGCTCC TATTTGGCAG 2040
AATTATTGTT AGAAAAAGGT TATGAAGTTC ATGGTATTAA ACGCCGTGCA TCTTCATTTA 2100
ATACTGAGCG AGTGGATCAC ATCTATCAGG ATTCACATTT AGCTAATCCT AAACTTTTTC 2160
TACACTATGG CGATTTGACA GATACTTCCA ATCTGACCCG TATTTTAAAA GAAGTTCAAC 2220

CAGATGAAGT TTACAATTTG GGGGCGATGA GCCATGTAGC GGTATCATTT GAGTCACCAG 2280
AATACACTGC TGATGTTGAT GCGATAGGAA CATTGCGTCT TCTTGAAGCT ATCAGGATAT 2340
TGGGGCTGGA AAAAAAGACA AAATTTTATC AGGCTTCAAC TTCAGAGCTT TATGGTTTGG 2400
TTCAAGAAAT TCCACAAAAA GAGACTACGC CATTTTATCC ACGTTCGCCT TATGCTGTTG 2460
CAAAATTATA TGCCTATTGG ATCACTGTTA ATTATCGTGA GTCTTATGGT ATGTTTGCCT 2520
GCAATGGTAT TCTCTTTAAC CACGAATCAC CTCGCCGTGG CGAGACCTTT GTTACTCGTA 2580
AAATAACACG CGGGATAGCA AATATTGCTC AAGGTCTTGA TAAATGCTTA TACTTGGGAA 2640
ATATGGATTC TCTGCGTGAT TGGGGACATG CTAAGGATTA TGTCAAAATG CAATGGATGA 2700
TGCTGCAGCA AGAAACTCCA GAAGATTTTG TAATTGCTAC AGGAATTCAA TATTCTGTCC 2760
GTGAGTTTGT CACAATGGCG GCAGAGCAAG TAGGCATAGA GTTAGCATTT GAAGGTGAGG 2820
GAGTAAATGA AAAAGGTGTT GTTGTTTCGG TCAATGGCAC TGATGCTAAA GCTGTAAACC 2880
CGGGCGATGT AATTATATCT GTAGATCCAA GGTATTTTAG GCCTGCAGAA GTTGAAACCT 2940
TGCTTGGCGA TCCTACTAAT GCGCATAAAA AATTAGGATG GAGCCCTGAA ATTACATTGC 3000
GTGAAATGGT AAAAGAAATG GTTTCCAGCG ATTTAGCAAT AGCGAAAAAG AACGTCTTGC 3060
TGAAAGCTAA TAACATTGCC ACTAATATTC CGCAAGAATA AAAAAGATAA TACATTAAAT 3120
AATTAAAAAT GGTGCTAGAT TTATTAGTAC CATTATTTTT TTTTGGGTGA CTAATGTTTA 3180
TTACATCAGA TAAATTTAGA GAAATTATCA AGTTAGTTCC ATTAGTATCA ATTGATCTGC 3240
TAATTGAAAA CGAGAATGGT GAATATTTAT TTGGTCTTAG GAATAATCGA CCGGCCAAAA 3300
ATTATTTTTT TGTTCCAGGT GGTAGGATTC GCAAAAATGA ATCTATTAAA AATGCTTTTA 3360
AAAGAATATC ATCTATGGAA TTAGGTAAAG AGTATGGTAT TTCAGGAAGT GTTTTTAATG 3420
GTGTATGGGA ACATTTCTAT GATGATGGTT TTTTTTCTGA AGGCGAGGCA ACACATTATA 3480
TAGTGCTTTG TTACACACTG AAAGTTCTTA AAAGTGAATT GAATCTCCCA GATGATCAAC 3540
ATCGTGAATA CCTTTGGCTA ACTAAACACC AAATAAATGC TAAACAAGAT GTTCATAACT 3600
ATTCAAAAAA TTATTTTTTG TAATTTTTAT TAAAAATTAA TATGCGAGAG AATTGTATGT 3660
CTCAATGTCT TTACCCTGTA ATTATTGCCG GAGGAACCGG AAGCCGTCTA TGGCCGTTGT 3720
CTCGAGTATT ATACCCTAAA CAATTTTTAA ATTTAGTTGG GGATTCTACA ATGTTGCAAA 3780
CAACAATTAC GCGTTTGGAT GGCATCGAAT GCGAAAATCC AATTGTTATC TGCAATGAAG 3840
ATCACCGATT TATTGTAGCA GAGCAATTAC GACAGATTGG TAAGCTAACC AAGAATATTA 3900
TACTTGAGCC GAAAGGCCGT AATACTGCAC CTGCCATAGC TTTAGCTGCT TTTATCGCTC 3960
AGAAGAATAA TCCTAATGAC GACCCTTTAT TATTAGTACT TGCGGCAGAC CACTCTATAA 4020
ATAATGAAAA AGCATTTCGA GAGTCAATAA TAAAAGCTAT GCCGTATGCA ACTTCTGGGA 4080
AGTTAGTAAC ATTTGGAATT ATTCCGGACA CGGCAAATAC TGGTTATGGA TATATTAAGA 4140
GAAGTTCTTC AGCTGATCCT AATAAAGAAT TCCCAGCATA TAATGTTGCG GAGTTTGTAG 4200
AAAAACCAGA TGTTAAAACA GCACAGGAAT ATATTTCGAG TGGGAATTAT TACTGGAATA 4260

GCGGAATGTT TTTATTTCGC GCCAGTAAAT ATCTTGATGA ACTACGGAAA TTTAGACCAG 4320
ATATTTATCA TAGCTGTGAA TGTGCAACCG CTACAGCAAA TATAGATATG GACTTTGTCC 4380
GAATTAACGA GGCTGAGTTT ATTAATTGTC CTGAAGAGTC TATCGATTAT GCTGTGATGG 4440
AAAAAACAAA AGACGCTGTA GTTCTTCCGA TAGATATTGG CTGGAATGAC GTGGGTTCTT 4500
GGTCATCACT TTGGGATATA AGCCAAAAGG ATTGCCATGG TAATGTGTGC CATGGGGATG 4560
TGCTCAATCA TGATGGAGAA AATAGTTTTA TTTACTCTGA GTCAAGTCTG GTTGCGACAG 4620
TCGGAGTAAG TAATTTAGTA ATTGTCCAAA CCAAGGATGC TGTACTGGTT GCGGACCGTG 4680
ATAAAGTCCA AAATGTTAAA AACATAGTTG ACGATCTAAA AAAGAGAAAA CGTGCTGAAT 4740
ACTACATGCA TCGTGCAGTT TTTCGCCCTT GGGGTAAATT CGATGCAATA GACCAAGGCG 4800
ATAGATATAG AGTAAAAAAA ATAATAGTTA AACCAGGAGA AGGGTTAGAT TTAAGGATGC 4860
ATCATCATAG GGCAGAGCAT TGGATTGTTG TATCCGGTAC TGCTAAAGTT TCACTAGGTA 4920
GTGAAGTTAA ACTATTAGTT TCTAATGAGT CTATATATAT CCCTCAGGGA GCAAAATATA 4980
GTCTTGAGAA TCCAGGCGTA ATACCTTTGC ATCTAATTGA AGTAAGTTCT GGTGATTACC 5040
TTGAATCAGA TGATATAGTG CGTTTTACTG ACAGATATAA CAGTAAACAA TTCCTAAAGC 5100
GAGATTGATA AATATGAATA AAATAACTTG CTTCAAAGCA TATGATATAC GTGGGCGTCT 5160
TGGTGCTGAA TTGAATGATG AAATAGCATA TAGAATTGGT CGCGCTTATG GTGAGTTTTT 5220
TAAACCTCAA ACTGTAGTTG TGGGAGGAGA TGCTCGCTTA ACAAGTGAGA GTTTAAAGAA 5280
ATCACTCTCA AATGGGCTAT GTGATGCAGG CGTAAATGTC TTAGATCTTG GAATGTGTGG 5340
TACTGAAGAG ATATATTTTT CCACTTGGTA TTTAGGAATT GATGGTGGAA TCGAGGTAAC 5400
TGCAAGCCAT AATCCAATTG ATTATAATGG AATGAAATTA GTAACCAAAG GTGCTCGACC 5460
AATCAGCAGT GACACAGGTC TCAAAGATAT ACAACAATTA GTAGAGAGTA ATAATTTTGA 5520
AGAGCTCAAC CTAGAAAAAA AAGGGAATAT TACCAAATAT TCCACCCGAG ATGCCTACAT 5580
AAATCATTTG ATGGGCTATG CTAATCTGCA AAAAATAAAA AAAATCAAAA TAGTTGTGAA 5640
TTCTGGGAAT GGTGCAGCTG GTCCTGTTAT TGATGCTATT GAGGAATGCT TTTTACGGAA 5700
CAATATTCCG ATTCAGTTTG TAAAAATAAA TAATACACCC GATGGTAATT TTCCACATGG 5760
TATCCCTAAT CCATTACTAC CTGAGTGCAG AGAAGATACC AGCAGTGCGG TTATAAGACA 5820
TAGTGCTGAT TTTGGTATTG CATTTGATGG TGATTTTGAT AGGTGTTTTT TCTTTGATGA 5880
AAATGGACAA TTTATTGAAG GATACTACAT TGTTGGTTTA TTAGCGGAAG TTTTTTTAGG 5940
GAAATATCCA AACGCAAAAA TCATTCATGA TCCTCGCCTT ATATGGAATA CTATTGATAT 6000
CGTAGAAAGT CATGGTGGTA TACCTATAAT GACTAAAACC GGTCATGCTT ACATTAAGCA 6060
AAGAATGCGT GAAGAGGATG CCGTATATGG CGGCGAAATG AGTGCGCATC ATTATTTTAA 6120
AGATTTTGCA TACTGCGATA GTGGAATGAT TCCTTGGATT TTAATTTGTG AACTTTTGAG 6180
TCTGACAAAT AAAAAATTAG GTGAACTGGT TTGTGGTTGT ATAAACGACT GGCCGGCAAG 6240
TGGAGAAATA AACTGTACAC TAGACAATCC GCAAAATGAA ATAGATAAAT TATTTAATCG 6300

TTACAAAGAT AGTGCCTTAG CTGTTGATTA CACTGATGGA TTAACTATGG AGTTCTCTGA 6360
TTGGCGTTTT AATGTTAGAT GCTCAAATAC AGAACCTGTA GTACGATTGA ATGTAGAATC 6420
TAGGAATAAT GCTATTCTTA TGCAGGAAAA AACAGAAGAA ATTCTGAATT TTATATCAAA 6480
ATAAATTTGC ACCTGAGTTC ATAATGGGAA CAAGAAATAT ATGAAAGTAC TTCTGACTGG 6540
CTCAACTGGC ATGGTTGGTA AGAATATATT AGAGCATGAT AGTGCAAGTA AATATAATAT 6600
ACTTACTCCA ACCAGCTCTG ATTTGAATTT ATTAGATAAA AATGAAATAG AAAAATTCAT 6660
GCTTATCAAC ATGCCAGACT GTATTATACA TGCAGCGGGA TTAGTTGGAG GCATTCATGC 6720
AAATATAAGC AGGCCGTTTG ATTTTCTGGA AAAAAATTTG CAGATGGGTT TAAATTTAGT 6780
TTCCGTCGCA AAAAAACTAG GTATCAAGAA AGTGCTTAAC TTGGGTAGTT CATGCATGTA 6840
CCCCAAAAAC TTTGAAGAGG CTATTCCTGA GAAAGCTCTG TTAACTGGTG AGCTAGAAGA 6900
AACTAATGAG GGATATGCTA TTGCGAAAAT TGCTGTAGCA AAAGCATGCG AATATATATC 6960
AAGAGAAAAC TCTAATTATT TTTATAAAAC AATTATCCCA TGTAATTTAT ATGGGAAATA 7020
TGATAAATTT GATGATAACT CGTCACATAT GATTCCGGCA GTTATAAAAA AAATCCATCA 7080
TGCGAAAATT AATAATGTCC CAGAGATCGA AATTTGGGGG GATGGTAATT CGCGCCGTGA 7140
GTTTATGTAT GCAGAAGATT TAGCTGATCT TATTTTTTAT GTTATTCCTA AAATAGAATT 7200
CATGCCTAAT ATGGTAAATG CTGGTTTAGG TTACGATTAT TCAATTAATG ACTATTATAA 7260
GATAATTGCA GAAGAAATTG GTTATACTGG GAGTTTTTCT CATGATTTAA CAAAACCAAC 7320
AGGAATGAAA CGGAAGCTAG TAGATATTTC ATTGCTTAAT AAAATTGGTT GGTCAAGTCA 7380
CTTTGAACTC AGAGATGGCA TCAGAAAGAC CTATAATTAT TACTTGGAGA ATCAAAATAA 7440
ATGATTACAT ACCCACTTGC TAGTAATACT TGGGATGAAT ATGAGTATGC AGCAATACAG 7500
TCAGTAATTG ACTCAAAAAT GTTTACCATG GGTAAAAAGG TTGAGTTATA TGAGAAAAAT 7560
TTTGCTGATT TGTTTGGTAG CAAATATGCC GTAATGGTTA GCTCTGGTTC TACAGCTAAT 7620
CTGTTAATGA TTGCTGCCCT TTTCTTCACT AATAAACCAA AACTTAAAAG AGGTGATGAA 7680
ATAATAGTAC CTGCAGTGTC ATGGTCTACG ACATATTACC CTCTGCAACA GTATGGCTTA 7740
AAGGTGAAGT TTGTCGATAT CAATAAAGAA ACTTTAAATA TTGATATCGA TAGTTTGAAA 7800
AATGCTATTT CAGATAAAAC AAAAGCAATA TTGACAGTAA ATTTATTAGG TAATCCTAAT 7860
GATTTTGCAA AAATAAATGA GATAATAAAT AATAGGGATA TTATCTTACT AGAAGATAAC 7920
TGTGAGTCGA TGGGCGCGGT CTTTCAAAAT AAGCAGGCAG GCACATTCGG AGTTATGGGT 7980
ACCTTTAGTT CTTTTTACTC TCATCATATA GCTACAATGG AAGGGGGCTG CGTAGTTACT 8040
GATGATGAAG AGCTGTATCA TGTATTGTTG TGCCTTCGAG CTCATGGTTG GACAAGAAAT 8100
TTACCAAAAG AGAATATGGT TACAGGCACT AAGAGTGATG ATATTTTCGA AGAGTCGTTT 8160
AAGTTTGTTT TACCAGGATA CAATGTTCGC CCACTTGAAA TGAGTGGTGC TATTGGGATA 8220
GAGCAACTTA AAAAGTTACC AGGTTTTATA TCCACCAGAC GTTCCAATGC ACAATATTTT 8280
GTAGATAAAT TTAAAGATCA TCCATTCCTT GATATACAAA AAGAAGTTGG TGAAAGTAGC 8340

TGGTTTGGTT TTTCCTTCGT TATAAAGGAG GGAGCTGCTA TTGAGAGGAA GAGTTTAGTA 8400
AATAATCTGA TCTCAGCAGG CATTGAATGC CGACCAATTG TTACTGGGAA TTTTCTCAAA 8460
AATGAACGTG TTTTGAGTTA TTTTGATTAC TCTGTACATG ATACGGTAGC AAATGCCGAA 8520
TATATAGATA AGAATGGTTT TTTTGTCGGA AACCACCAGA TACCTTTGTT TAATGAAATA 8580
GATTATCTAC GAAAAGTATT AAAATAACTA ACGAGGCACT CTATTTCGAA TAGAGTGCCT 8640
TTAAGATGGT ATTAACAGTG AAAAAAATTT TAGCGTTTGG CTATTCTAAA GTACTACCAC 8700
CGGTTATTGA ACAGTTTGTC AATCCAATTT GCATCTTCAT TATCACACCA CTAATACTCA 8760
ACCACCTGGG TAAGCAAAGC TATGGTAATT GGATTTTATT AATTACTATT GTATCTTTTT 8820
CTCAGTTAAT ATGTGGAGGA TGTTCCGCAT GGATTGCAAA AATCATTGCA GAACAGAGAA 8880
TTCTTAGTGA TTTATCAAAA AAAAATGCTT TACGTCAAAT TTCCTATAAT TTTTCAATTG 8940
TTATTATCGC ATTTGCGGTA TTGATTTCTT TTCTTATATT AAGTATTTGT TTCTTCGATG 9000
TTGCGAGGAA TAATTCTTCA TTCTTATTCG CGATTATTAT TTGTGGTTTT TTTCAGGAAG 9060
TTGATAATTT ATTTAGTGGT GCGCTAAAAG GTTTTGAAAA ATTTAATGTA TCATGTTTTT 9120
TTGAAGTAAT TACAAGAGTG CTCTGGGCTT CTATAGTAAT ATATGGCATT TACGGAAATG 9180
CACTCTTATA TTTTACATGT TTAGCCTTTA CCATTAAAGG TATGCTAAAA TATATTCTTG 9240
TATGTCTGAA TATTACCGGT TGTTTCATCA ATCCTAATTT TAATAGAGTT GGGATTGTTA 9300
ATTTGTTAAA TGAGTCAAAA TGGATGTTTC TTCAATTAAC TGGTGGCGTC TCACTTAGTT 9360
TGTTTGATAG GCTCGTAATA CCATTGATTT TATCTGTCAG TAAACTGGCT TCTTATGTCC 9420
CTTGCCTTCA ACTAGCTCAA TTGATGTTCA CTCTTTCTGC GTCTGCAAAT CAAATATTAC 9480
TACCAATGTT TGCTAGAATG AAAGCATCTA ACACATTTCC CTCTAATTGT TTTTTTAAAA 9540
TTCTGCTTGT ATCACTAATT TCTGTTTTGC CTTGTCTTGC GTTATTCTTT TTTGGTCGTG 9600
ATATATTATC AATATGGATA AACCCTACAT TTGCAACTGA AAATTATAAA TTAATGCAAA 9660
TTTTAGCTAT AAGTTACATT TTATTGTCAA TGATGACATC TTTTCATTTC TTGTTATTAG 9720
GAATTGGTAA ATCTAAGCTT GTTGCAAATT TAAATCTGGT TGCAGGGCTC GCACTTGCTG 9780
CTTCAACGTT AATCGCAGCT CATTATGGCC TTTATGCAAT ATCTATGGTA AAAATAATAT 9840
ATCCGGCTTT TCAATTTTAT TACCTTTATG TAGCTTTTGT CTATTTTAAT AGAGCGAAAA 9900
ATGTCTATTG ATTTACTTTT TTCAATTACT GAAATCGCAA TTGTTTTTTC TTGCACTATT 9960
TACATATTTA CTCAATGTTT GTTAATGCGG AGGATCTATT TAGATAAAAG TATTTTAATT 10020
CTTTTATGCT TGCTCTTTTT TTTAGTAATC ATTCAACTTC CTGAGCTTAA TGTAAACGGT 10080
TTGGTCGATT CTTTAAAGTT ATCACTGCCT TTATTGATGG TCTTTATCGC TTTTCAAAAA 10140
CCGAAATTAT GCTTGTGGGT TATTATTGCA TTGTTGTTTT TGAACTCTGC ATTTAATTTT 10200
TTATATTTAA AGACATTCGA TAAGTTTAGC TCATTTCCTT TTACTTTTTT TATATTGCTG 10260
TTTTACTTGT TTAGATTGGG AATTGGTAAT TTACCGGTTT ATAAAAATAA AAAATTTTAC 10320
GCGTTGATTT TTCTCTTTAT ATTAATAGAC ATAATGCAGT CATTGTTAAT AAATTATAGG 10380

GGGCAGATTT TATATTCCGT AATTTGCATC CTGATACTTG TGTTTAAAGT TAATTTAAGA 10440
AAAAAGATTC CATACTTTTT TTTAATGCTG CCAGTTTTAT ATGTAATTAT TATGGCTTAT 10500
ATTGGTTTTA ATTATTTCAA TAAAGGCGTA ACTTTTTTTG AACCTACAGC AAGTAATATT 10560
GAACGTACGG GGATGATATA TTATTTGGTT TCACAGCTTG GTGATTATAT ATTCCATGGT 10620
ATGGGGACAT TAAATTTCTT AAATAACGGC GGACAATATA AGACGTTATA TGGACTTCCA 10680
TCATTAATTC CTAATGACCC TCATGATTTT TTATTACGGT TCTTTATAAG TATTGGTGTG 10740
ATAGGAGCAT TGGTTTATCA TTCTATATTT TTTGTTTTTT TTAGGAGAAT ATCTTTCTTA 10800
TTATATGAGA GAAATGCTCC TTTCATTGTT GTAAGTTGTT TGTTACTGTT ACAAGTTGTG 10860
TTAATTTATA CATTAAACCC TTTTGATGCT TTTAATCGAT TGATTTGCGG GCTTACAGTT 10920
GGAGTTGTTT ATGGATTTGC AAAAATTAGA TAAGTATACC TGTAATGGAA ATTTAGACGC 10980
TCCACTTGTT TCAATAATCA TTGCAACTTA TAATTCTGAA CTTGATATAG CTAAGTGTTT 11040
GCAATCGGTA ACTAATCAAT CTTATAAGAA TATTGAAATC ATAATAATGG ATGGAGGATC 11100
TTCTGATAAA ACGCTTGATA TTGCAAAATC GTTTAAAGAC GACCGAATAA AAATAGTTTC 11160
AGAGAAAGAT CGTGGAATTT ATGATGCCTG GAATAAAGCA GTTGATTTAT CCATTGGTGA 11220
TTGGGTAGCA TTTATTGGTT CAGATGATGT TTACTATCAT ACAGATGCAA TTGCTTCATT 11280
GATGAAGGGG GTTATGGTAT CTAATGGCGC CCCTGTGGTT TATGGGAGGA CAGCGCACGA 11340
AGGTCCCGAT AGGAACATAT CTGGATTTTC AGGCAGTGAA TGGTACAACC TAACAGGATT 11400
TAAGTTTAAT TATTACAAAT GTAATTTACC ATTGCCCATT ATGAGCGCAA TATATTCTCG 11460
TGATTTCTTC AGAAACGAAC GTTTTGATAT TAAATTAAAA ATTGTTGCTG ACGCTGATTG 11520
GTTTCTGAGA TGTTTCATCA AATGGAGTAA AGAGAAGTCA CCTTATTTTA TTAATGACAC 11580
GACCCCTATT GTTAGAATGG GATATGGTGG GGTTTCGACT GATATTTCTT CTCAAGTTAA 11640
AACTACGCTA GAAAGTTTCA TTGTACGCAA AAAGAATAAT ATATCCTGTT TAAACATACA 11700
GCTGATTCTT AGATATGCTA AAATTCTGGT GATGGTAGCG ATCAAAAATA TTTTTGGCAA 11760
TAATGTTTAT AAATTAATGC ATAACGGGTA TCATTCCCTA AAGAAAATCA AGAATAAAAT 11820
ATGAAGATTG TTTATATAAT AACCGGGCTT ACTTGTGGTG GAGCCGAACA CCTTATGACG 11880
CAGTTAGCAG ACCAAATGTT TATACGCGGG CATGATGTTA ATATTATTTG TCTAACTGGT 11940
ATATCTGAGG TAAAGCCAAC ACAAAATATT AATATTCATT ATGTTAATAT GGATAAAAAT 12000
TTTAGAAGCT TTTTTAGAGC TTTATTTCAA GTAAAAAAAA TAATTGTCGC CTTAAAGCCA 12060
GATATAATAC ATAGTCATAT GTTTCATGCT AATATTTTTA GTCGTTTTAT TAGGATGCTG 12120
ATTCCAGCGG TGCCCCTGAT ATGTACCGCA CACAACAAAA ATGAAGGTGG CAATGCAAGG 12180
ATGTTTTGTT ATCGACTGAG TGATTTTTTA GCTTCTATTA CTACAAATGT AAGTAAAGAG 12240
GCTGTTCAAG AGTTTATAGC AAGAAAGGCT ACACCTAAAA ATAAAATAGT AGAGATTCCG 12300
AATTTTATTA ATACAAATAA ATTTGATTTT GATATTAATG TCAGAAAGAA AACGCGAGAT 12360
GCTTTTAATT TGAAAGACAG TACAGCAGTA CTGCTCGCAG TAGGAAGACT TGTTGAAGCA 12420

AAAGACTATC CGAACTTATT AAATGCAATA AATCATTTGA TTCTTTCAAA AACATCAAAT 12480
TGTAATGATT TTATTTTGCT TATTGCTGGC GATGGCGCAT TAAGAAATAA ATTATTGGAT 12540
TTGGTTTGTC AATTGAATCT TGTGGATAAA GTTTTCTTCT TGGGGCAAAG AAGTGATATT 12600
AAAGAATTAA TGTGTGCTGC AGATCTTTTT GTTTTGAGTT CTGAGTGGGA AGGTTTTGGT 12660
CTCGTTGTTG CAGAAGCTAT GGCGTGTGAA CGTCCCGTTG TTGCTACCGA TTCTGGTGGA 12720
GTTAAAGAAG TCGTTGGACC TCATAATGAT GTTATCCCTG TCAGTAATCA TATTCTGTTG 12780
GCAGAGAAAA TCGCTGAGAC ACTTAAAATA GATGATAACG CAAGAAAAAT AATAGGTATG 12840
AAAAATAGAG AATATATTGT TTCCAATTTT TCAATTAAAA CGATAGTGAG TGAGTGGGAG 12900
CGCTTATATT TTAAATATTC CAAGCGTAAT AATATAATTG ATTGAAAATA TAAGTTTGTA 12960
CTCTGGATGC AATAGTTTCT CTATGCTGTT TTTTTACTGG CTCCGTATTT TTACTTATAG 13020
CTGGATTTTG TTATATATCA GTATTAATCT GTCTCAACTT CATCTAGACT ACATTCAAGC 13080
CGCGCATGCG TCGCGCGGTG ACTACACCTG ACAGGAGTAT GTAATGTCCA AGCAACAGAT 13140
CGGCGTCGTC GGTATGGCAG TGATGGGGCG CAACCTGGCG CTCAACATCG AAAGCCGCGG 13200
TTATACCGTC TCCATCTTCA ACCGCTCCCG CGAGAAAACT GAAGAAGTTG TTGCCGAGAA 13260
CCCGGATAAG AAACTGGTTC CTTATTACAC GGTGAAAGAG TTCGTCGAGT CTCTTGAAAC 13320
CCCACGTCGT ATCCTGTTAA TGGTAAAAGC AGGGGCGGGA ACTGATGCTG CTATCGATTC 13380
CCTGAAGCCG TATCTGGATA AAGGCGACAT CATTATTGAT GGTGGCAACA CCTTCTTCCA 13440
GGACACTATC CGTCGTAACC GTGAACTGTC CGCGGAAGGC TTTAACTTCA TCGGTACCGG 13500
CGTGTCCGGC GGTGAAGAGG GCGCCCTGAA AGGCCCATCT ATCATGCCAG GTGGCCAGAA 13560
AGAAGCGTAT GAGCTGGTTG CGCCTATCCT GACCAAGATT GCTGCGGTTG CTGAAGATGG 13620
CGAACCATGT ATAACTTACA TCGGTGCTGA CGGTGCGGGT CACTACGTGA AGATGGTGCA 13680
CAACGGTATC GAATATGGCG ATATGCAGCT GATTGCTGAA GCCTATTCTC TGCTTAAAGG 13740
CGGCCTTAAT CTGTCTAACG AAGAGCTGGC AACCACTTTT ACCGAGTGGA ATGAAGGCGA 13800
GCTAAGTAGC TACCTGATTG ACATCACCAA AGACATCTTC ACCAAAAAAG ATGAAGAGGG 13860
TAAATACCTG GTTGATGTGA TCCTGGACGA AGCTGCGAAC AAAGGCACCG GTAAATGGAC 13920
CAGCCAGAGC TCTCTGGATC TGGGTGAACC GCTGTCGCTG ATCACCGAAT CCGTATTCGC 13980
TCGCTACATC TCTTCTCTGA AAGACCAGCG CATTGCGGCA TCTAAAGTGC TGTCTGGTCC 14040
GCAGGCTAAA CTGGCTGGTG ATAAAGCAGA GTTCGTTGAG AAAGTCCGTC GCGCGCTGTA 14100
CCTGGGTAAA ATCGTCTCTT ATGCCCAAGG CTTCTCTCAA CTGCGTGCCG CGTCTGACGA 14160
ATACAACTGG GATCTGAACT ACGGCGAAAT CGCGAAGATC TTCCGCGCGG GCTGCATCAT 14220
TCGTGCGCAG TTCCTGCAGA AAATTACTGA CGCGTATGCT GAAAACAAAG GCATTGCTAA 14280
CCTGTTGCTG GCTCCGTACT TCAAAAATAT CGCTGATGAA TATCAGCAAG CGCTGCGTGA 14340
TGTAGTGGCT TATGCTGTGC AGAACGGTAT TCCGGTACCG ACCTTCTCTG CAGCGGTAGC 14400
TACTACGAC AGCTACCGTT CTGCGGTACT GCCGGCTAAT CTGATTCAGG CACAGCGTGA 14460

TTACTTCGGT GCGCACACGT ATAAACGCAC TGATAAAGAA GGTGTGTTCC ACACCG 14516

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14024 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(v) ORIGINAL SOURCE

(A) ORGANISM: Escherichia coli

(vi) Note that the first 19bp is from the primer used for the long PCR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GTAACCAAGG GCGGTACGTG CATAAATTTT AATGCTTATC AAAACTATTA GCATTAAAAA 60

TATATAAGAA ATTCTCAAAT GAACAAAGAA ACCGTTTCAA TAATTATGCC CGTTTACAAT 120

GGGGCCAAAA CTATAATCTC ATCAGTAGAA TCAATTATAC ATCAATCTTA TCAAGATTTT 180

GTTTTGTATA TCATTGACGA TTGTAGCACC GATGATACAT TTTCATTAAT CAACAGTCGA 240

TACAAAAACA ATCAGAAAAT AAGAATATTG CGTAACAAGA CAAATTTAGG TGTTGCAGAA 300

AGTCGAAATT ATGGAATAGA AATGGCCACG GGGAAATATA TTTCTTTTTG TGATGCGGAT 360

GATTTGTGGC ACGAGAAAAA ATTAGAGCGT CAAATCGAAG TGTTAAATAA TGAATGTGTA 420

GATGTGGTAT GTTCTAATTA TTATGTTATA GATAACAATA GAAATATTGT TGGCGAAGTT 480

AATGCTCCTC ATGTGATAAA TTATAGAAAA ATGCTCATGA AAAACTACAT AGGGAATTTG 540

ACAGGAATCT ATAATGCCAA CAAATTGGGT AAGTTTTATC AAAAAAAGAT TGGTCACGAG 600

GATTATTTGA TGTGGCTGGA AATAATTAAT AAAACAAATG GTGCTATTTG TATTCAAGAT 660

AATCTGGCGT ATTACATGCG TTCAAATAAT TCACTATCGG GTAATAAAAT TAAAGCTGCA 720

AAATGGACAT GGAGTATATA TAGAGAACAT TTACATTTGT CCTTTCCAAA AACATTATAT 780

TATTTTTTAT TATATGCTTC AAATGGAGTC ATGAAAAAAA TAACACATTC ACTATTAAGG 840

AGAAAGGAGA CTAAAAAGTG AAGTCAGCGG CTAAGTTGAT TTTTTTATTC CTATTTACAC 900

TTTATAGTCT CCAGTTGTAT GGGGTTATCA TAGATGATCG TATAACAAAT TTTGATACAA 960

AGGTATTAAC TAGTATTATA ATTATATTTC AGATTTTTTT TGTTTTATTA TTTTATCTAA 1020

CGATTATAAA TGAAAGAAAA CAGCAGAAAA AATTTATCGT GAACTGGGAG CTAAAGTTAA 1080

TACTCGTTTT CCTTTTTGTG ACTATAGAAA TTGCTGCTGT AGTTTTATTT CTTAAAGAAG 1140

GTATTCCTAT ATTTGATGAT GATCCAGGGG GGGCTAAACT TAGAATAGCT GAAGGTAATG 1200

GACTTTACAT TAGATATATT AAGTATTTTG GTAATATAGT TGTGTTTGCA TTAATTATTC 1260

TTTATGATGA GCATAAATTC AAACAGAGGA CCATCATATT TGTATATTTT ACAACGATTG 1320

CTTTATTTGG TTATCGTTCT GAATTGGTGT TGCTCATTCT TCAATATATA TTGATTACCA 1380

ATATCCTGTC AAAGGATAAC CGTAATCCTA AAATAAAAAG AATAATAGGG TATTTTTTAT 1440

TGGTAGGGGT TGTATGCTCG TTGTTTTATC TAAGTTTAGG ACAAGACGGA GAACAAAATG 1500

ACTCATATAA TAATATGTTA AGGATAATTA ATAGGTTAAC AATAGAGCAA GTTGAAGGTG 1560

TTCCATATGT TGTTTCTGAA TCTATTAAGA ACGATTTCTT TCCGACACCA GAGTTAGAAA 1620

AGGAATTAAA AGCAATAATA AATAGAATAC AGGGAATAAA GCATCAAGAC TTATTTTATG 1680

GAGAACGGTT ACATAAACAA GTATTTGGAG ACATGGGAGC AAATTTTTTA TCAGTTACTA 1740

CGTATGGAGC AGAACTGTTA GTTTTTTTTG GTTTTCTCTG TGTATTCATT ATCCCTTTAG 1800

GGATATATAT ACCTTTTTAT CTTTTAAAGA GAATGAAAAA AACCCATAGC TCGATAAATT 1860

GCGCATTCTA TTCATATATC ATTATGATTT TATTGCAATA CTTAGTGGCT GGGAATGCAT 1920

CGGCCTTCTT TTTTGGTCCT TTTCTCTCCG TATTGATAAT GTGTACTCCT CTGATCTTAT 1980

TGCATGATAC GTTAAAGAGA TTATCACGAA ATGAAAATAT CAGTTATAAC TGTGACTTAT 2040

AATAATGCTG AAGGGTTAGA AAAAACTTTA AGTAGTTTAT CAATTTTAAA AATAAAACCT 2100

TTTGAGATTA TTATAGTTGA TGGCGGCTCT ACAGATGGAA CGAATCGTGT CATTAGTAGA 2160

TTTACTAGTA TGAATATTAC ACATGTTTAT GAAAAAGATG AAGGGATATA TGATGCGATG 2220

AATAAGGGCC GAATGTTGGC CAAAGGCGAC TTAATACATT ATTTAAACGC CGGCGATAGC 2280

GTAATTGGAG ATATATATAA AAATATCAAA GAGCCATGTT TGATTAAAGT TGGCCTTTTC 2340

GAAAATGATA AACTTCTGGG ATTTTCTTCT ATAACCCATT CAAATACAGG GTATTGTCAT 2400

CAAGGGGTGA TTTTCCCAAA GAATCATTCA GAATATGATC TAAGGTATAA AATATGTGCT 2460

GATTATAAGC TTATTCAAGA GGTGTTTCCT GAAGGGTTAA GATCTCTATC TTTGATTACT 2520

TCGGGTTATG TAAAATATGA TATGGGGGGA GTATCTTCAA AAAAAAGAAT TTTAAGAGAT 2580

AAAGAGCTTG CCAAAATTAT GTTTGAAAAA AATAAAAAAA ACCTTATTAA GTTTATTCCA 2640

ATTTCAATAA TCAAAATTTT ATTCCCTGAA CGTTTAAGAA GAGTATTGCG GAAAATGCAA 2700

TATATTTGTC TAACTTTATT CTTCATGAAG AATAGTTCAC CATATGATAA TGAATAAAAT 2760

CAAAAAAATA CTTAAATTTT GCACTTTAAA AAAATATGAT ACATCAAGTG CTTTAGGTAG 2820

AGAACAGGAA AGGTACAGGA TTATATCCTT GTCTGTTATT TCAAGTTTGA TTAGTAAAAT 2880

ACTCTCACTA CTTTCTCTTA TATTAACTGT AAGTTTAACT TTACCTTATT TAGGACAAGA 2940

GAGATTTGGT GTATGGATGA CTATTACCAG TCTTGGTGCT GCTCTGACAT TTTTGGACTT 3000

AGGTATAGGA AATGCATTAA CAAACAGGAT CGCACATTCA TTTGCGTGTG GCAAAAATTT 3060

AAAGATGAGT CGGCAAATTA GTGGTGGGCT CACTTTGCTG GCTGGATTAT CGTTTGTCAT 3120

AACTGCAATA TGCTATATTA CTTCTGGCAT GATTGATTGG CAACTAGTAA TAAAAGGTAT 3180

AAACGAGAAT GTGTATGCAG AGTTACAACA CTCAATTAAA GTCTTTGTAA TCATATTTGG 3240

ACTTGGAATT TATTCAAATG GTGTGCAAAA AGTTTATATG GGAATACAAA AAGCCTATAT 3300

AAGTAATATT GTTAATGCCA TATTTATATT GTTATCTATT ATTACTCTAG TAATATCGTC 3360

GAAACTACAT GCGGGACTAC CAGTTTTAAT TGTCAGCACT CTTGGTATTC AATACATATC 3420

GGGAATCTAT TTAACAATTA ATCTTATTAT AAAGCGATTA ATAAAGTTTA CAAAAGTTAA 3480

CATACATGCT AAAAGAGAAG CTCCATATTT GATATTAAAC GGTTTTTTCT TTTTTATTTT 3540

ACAGTTAGGC ACTCTGGCAA CATGGAGTGG TGATAACTTT ATAATATCTA TAACATTGGG 3600

TGTTACTTAT GTTGCTGTTT TTAGCATTAC ACAGAGATTA TTTCAAATAT CTACGGTCCC 3660

TCTTACGATT TATAACATCC CGTTATGGGC TGCTTATGCA GATGCTCATG CACGCAATGA 3720

TACTCAATTT ATAAAAAAGA CGCTCAGAAC ATCATTGAAA ATAGTGGGTA TTTCATCATT 3780

CTTATTGGCC TTCATATTAG TAGTGTTCGG TAGTGAAGTC GTTAATATTT GGACAGAAGG 3840

AAAGATTCAG GTACCTCGAA CATTCATAAT AGCTTATGCT TTATGGTCTG TTATTGATGC 3900

TTTTTCGAAT ACATTTGCAA GCTTTTTAAA TGGTTTGAAC ATAGTTAAAC AACAAATGCT 3960

TGCTGTTGTA ACATTGATAT TGATCGCAAT TCCAGCAAAA TACATCATAG TTAGCCATTT 4020

TGGGTTAACT GTTATGTTGT ACTGCTTCAT TTTTATATAT ATTGTAAATT ACTTTATATG 4080

GTATAAATGT AGTTTTAAAA AACATATCGA TAGACAGTTA AATATAAGAG GATGAAAATG 4140

AAATATATAC CAGTTTACCA ACCGTCATTG ACAGGAAAAG AAAAAGAATA TGTAAATGAA 4200

TGTCTGGACT CAACGTGGAT TTCATCAAAA GGAAACTATA TTCAGAAGTT TGAAAATAAA 4260

TTTGCGGAAC AAAACCATGT GCAATATGCA ACTACTGTAA GTAATGGAAC GGTTGCTCTT 4320

CATTTAGCTT TGTTAGCGTT AGGTATATCG GAAGGAGATG AAGTTATTGT TCCAACACTG 4380

ACATATATAG CATCAGTTAA TGCTATAAAA TACACAGGAG CCACCCCCAT TTTCGTTGAT 4440

TCAGATAATG AAACTTGGCA AATGTCTGTT AGTGACATAG AACAAAAAAT CACTAATAAA 4500

ACTAAAGCTA TTATGTGTGT CCATTTATAC GGACATCCAT GTGATATGGA ACAAATTGTA 4560

GAACTGGCCA AAAGTAGAAA TTTGTTTGTA ATTGAAGATT GCGCTGAAGC CTTTGGTTCT 4620

AAATATAAAG GTAAATATGT GGGAACATTT GGAGATATTT CTACTTTTAG CTTTTTTGGA 4680

AATAAAACTA TTACTACAGG TGAAGGTGGA ATGGTTGTCA CGAATGACAA AACACTTTAT 4740

GACCGTTGTT TACATTTTAA AGGCCAAGGA TTAGCTGTAC ATAGGCAATA TTGGCATGAC 4800

GTTATAGGCT ACAATTATAG GATGACAAAT ATCTGCGCTG CTATAGGATT AGCCCAGTTA 4860

GAACAAGCTG ATGATTTTAT ATCACGAAAA CGTGAAATTG CTGATATTTA TAAAAAAAAT 4920

ATCAACAGTC TTGTACAAGT CCACAAGGAA AGTAAAGATG TTTTTCACAC TTATTGGATG 4980

GTCTCAATTC TAACTAGGAC CGCAGAGGAA AGAGAGGAAT TAAGGAATCA CCTTGCAGAT 5040

AAACTCATCG AAACAAGGCC AGTTTTTTAC CCTGTCCACA CGATGCCAAT GTACTCGGAA 5100

AAATATCAAA AGCACCCTAT AGCTGAGGAT CTTGGTTGGC GTGGAATTAA TTTACCTAGT 5160

TTCCCCAGCC TATCGAATGA GCAAGTTATT TATATTTGTG AATCTATTAA CGAATTTTAT 5220

AGTGATAAAT AGCCTAAAAT ATTGTAAAGG TCATTCATGA AAATTGCGTT GAATTCAGAT 5280

GGATTTTACG AGTGGGGCGG TGGAATTGAT TTTATTAAAT ATATTCTGTC AATATTAGAA 5340

ACGAAACCAG AAATATGTAT CGATATTCTT TTACCGAGAA ATGATATACA TTCTCTTATA 5400

AGAGAAAAAG CATTTCCTTT TAAAAGTATA TTAAAAGCAA TTTTAAAGAG GGAAAGGCCT 5460

CGATGGATTT CATTAAATAG ATTTAATGAG CAATACTATA GAGATGCCTT TACACAAAAT 5520

AATATAGAGA CGAATCTTAC CTTTATTAAA AGTAAGAGCT CTGCCTTTTA TTCATATTTT 5580

GATAGTAGCG ATTGTGATGT TATTCTTCCT TGCATGCGTG TTCCTTCGGG AAATTTGAAT 5640

AAAAAAGCAT GGATTGGTTA TATTTATGAC TTTCAACACT GTTACTATCC TTCATTTTTT 5700

AGTAAGCGAG AAATAGATCA AAGGAATGTG TTTTTTAAAT TGATGCTCAA TTGCGCTAAC 5760

AATATTATTG TTAATGCACA TTCAGTTATT ACCGATGCAA ATAAATATGT TGGGAATTAT 5820

TCTGCAAAAC TACATTCTCT TCCATTTAGT CCATGCCCTC AATTAAAATG GTTCGCTGAT 5880

TACTCTGGTA ATATTGCCAA ATATAATATT GACAAGGATT ATTTTATAAT TTGCAATCAA 5940

TTTTGGAAAC ATAAAGATCA TGCAACTGCT TTTAGGGCAT TTAAAATTTA TACTGAATAT 6000

AATCCTGATG TTTATTTAGT ATGCACGGGA GCTACTCAAG ATTATCGATT CCCTGGATAT 6060

TTTAATGAAT TGATGGTTTT GGCAAAAAAG CTCGGAATTG AATCGAAAAT TAAGATATTA 6120

GGGCATATAC CTAAACTTGA ACAAATTGAA TTAATCAAAA ATTGCATTGC TGTAATACAA 6180

CCAACCTTAT TTGAAGGCGG GCCTGGAGGG GGGGTAACAT TTGACGCTAT TGCATTAGGG 6240

AAAAAAGTTA TACTATCTGA CATAGATGTC AATAAAGAAG TTAATTGCGG TGATGTATAT 6300

TTCTTTCAGG CAAAAAACCA TTATTCATTA AATGACGCGA TGGTAAAAGC TGATGAATCT 6360

AAAATTTTTT ATGAACCTAC AACTCTGATA GAATTGGGTC TCAAAAGACG CAATGCGTGT 6420

GCAGATTTTC TTTTAGATGT TGTGAAACAA GAAATTGAAT CCCGATCTTA ATATATTCAA 6480

GAGGTATATA ATGACTAAAG TCGCTCTTAT TACAGGTGTA ACTGGACAAG ATGGATCTTA 6540

TCTAGCTGAG TTTTTGCTTG ATAAAGGGTA TGAAGTTCAT GGTATCAAAC GCCGAGCCTC 6600

ATCTTTTAAT ACAGAACGCA TAGACCATAT TTATCAAGAT CCACATGGTT CTAACCCAAA 6660

TTTTCACTTG CACTATGGAG ATCTGACTGA TTCATCTAAC CTCACTAGAA TTCTAAAGGA 6720

GGTACAGCCA GATGAAGTAT ATAATTTAGC TGCTATGAGT CACGTAGCAG TTTCTTTTGA 6780

GTCTCCAGAA TATACAGCCG ATGTCGATGC AATTGGTACA TTACGTTTAC TGGAAGCAAT 6840

TCGCTTTTTA GGATTGGAAA ACAAAACGCG TTTCTATCAA GCTTCAACCT CAGAATTATA 6900

TGGACTTGTT CAGGAAATCC CTCAAAAAGA ATCCACCCCT TTTTATCCTC GTTCCCCTTA 6960

TGCAGTTGCA AAACTTTACG CATATTGGAT CACGGTAAAT TATCGAGAGT CATATGGTAT 7020

TTATGCATGT AATGGTATAT TGTTCAATCA TGAATCTCCA CGCCGTGGAG AAACGTTTGT 7080

AACAAGGAAA ATTACTCGAG GACTTGCAAA TATTGCACAA GGCTTGGAAT CATGTTTGTA 7140

TTTAGGGAAT ATGGATTCGT TACGAGATTG GGGACATGCA AAAGATTATG TTAGAATGCA 7200

ATGGTTGATG TTACAACAGG AGCAACCCGA AGATTTTGTG ATTGCAACAG GAGTCCAATA 7260

CTCAGTCCGT CAGTTTGTCG AAATGGCAGC AGCACAACTT GGTATTAAGA TGAGCTTTGT 7320

TGGTAAAGGA ATCGAAGAAA AAGGCATTGT AGATTCGGTT GAAGGACAGG ATGCTCCAGG 7380

TGTGAAACCA GGTGATGTCA TTGTTGCTGT TGATCCTCGT TATTTCCGAC CAGCTGAAGT 7440

TGATACTTTG CTTGGAGATC CGAGCAAAGC TAATCTCAAA CTTGGTTGGA GACCAGAAAT 7500

TACTCTTGCT GAAATGATTT CTGAAATGGT TGCCAAAGAT CTTGAAGCCG CTAAAAAACA 7560

TTCTCTTTTA AAATCGCATG GTTTTTCTGT AAGCTTAGCT CTGGAATGAT GATGAATAAG 7620

CAACGTATTT TTATTGCTGG TCACCAAGGA ATGGTTGGAT CAGCTATTAC CCGACGCCTC 7680

AAACAACGTG ATGATGTTGA GTTGGTTTTA CGTACTCGGG ATGAATTGAA CTTGTTGGAT 7740

AGTAGCGCTG TTTTGGATTT TTTTTCTTCA CAGAAAATCG ACCAGGTTTA TTTGGCAGCA 7800

GCAAAAGTCG GAGGTATTTT AGCTAACAGT TCTTATCCTG CCGATTTTAT ATATGAGAAT 7860

ATAATGATAG AGGCGAATGT CATTCATGCT GCCCACAAAA ATAATGTAAA TAAACTGCTT 7920

TTCCTCGGTT CGTCGTGTAT TTATCCTAAG TTAGCACACC AACCGATTAT GGAAGACGAA 7980

TTATTACAAG GGAAACTTGA GCCAACAAAT GAACCTTATG CTATCGCAAA AATTGCAGGT 8040

ATTAAATTAT GTGAATCTTA TAACCGTCAG TTTGGGCGTG ATTACCGTTC AGTAATGCCA 8100

ACCAATCTTT ATGGTCCAAA TGACAATTTT CATCCAAGTA ATTCTCATGT GATTCCGGCG 8160

CTTTTGCGCC GCTTTCATGA TGCTGTGGAA AACAATTCTC CGAATGTTGT TGTTTGGGGA 8220

AGTGGTACTC CAAAGCGTGA ATTCTTACAT GTAGATGATA TGGCTTCTGC AAGCATTTAT 8280

GTCATGGAGA TGCCATACGA TATATGGCAA AAAAATACTA AAGTAATGTT GTCTCATATC 8340

AATATTGGAA CAGGTATTGA CTGCACGATT TGTGAGCTTG CGGAAACAAT AGCAAAAGTT 8400

GTAGGTTATA AAGGGCATAT TACGTTCGAT ACAACAAAGC CCGATGGAGC CCCTCGAAAA 8460

CTACTTGATG TAACGCTTCT TCATCAACTA GGTTGGAATC ATAAAATTAC CCTTCACAAG 8520

GGTCTTGAAA ATACATACAA CTGGTTTCTT GAAAACCAAC TTCAATATCG GGGGTAATAA 8580

TGTTTTTACA TTCCCAAGAC TTTGCCACAA TTGTAAGGTC TACTCCTCTT ATTTCTATAG 8640

ATTTGATTGT GGAAAACGAG TTTGGCGAAA TTTTGCTAGG AAAACGAATC AACCGCCCGG 8700

CACAGGGCTA TTGGTTCGTT CCTGGTGGTA GGGTGTTGAA AGATGAAAAA TTGCAGACAG 8760

CCTTTGAACG ATTGACAGAA ATTGAACTAG GAATTCGTTT GCCTCTCTCT GTGGGTAAGT 8820

TTTATGGTAT CTGGCAGCAC TTCTACGAAG ACAATAGTAT GGGGGGAGAC TTTTCAACGC 8880

ATTATATAGT TATAGCATTC CTTCTTAAAT TACAACCAAA CATTTTGAAA TTACCGAAGT 8940

CACAACATAA TGCTTATTGC TGGCTATCGC GAGCAAAGCT GATAAATGAT GACGATGTGC 9000

ATTATAATTG TCGCGCATAT TTTAACAATA AAACAAATGA TGCGATTGGC TTAGATAATA 9060

AGGATATAAT ATGTCTGATG CGCCAATAAT TGCTGTAGTT ATGGCCGGTG GTACAGGCAG 9120

TCGTCTTTGG CCACTTTCTC GTGAACTATA TCCAAAGCAG TTTTTACAAC TCTCTGGTGA 9180

TAACACCTTG TTACAAACGA CTTTGCTACG ACTTTCAGGC CTATCATGTC AAAAACCATT 9240

AGTGATAACA AATGAACAGC ATCGCTTTGT TGTGGCTGAA CAGTTAAGGG AAATAAATAA 9300

ATTAAATGGT AATATTATTC TAGAACCATG CGGGCGAAAT ACTGCACCAG CAATAGCGAT 9360

ATCTGCGTTT CATGCGTTAA AACGTAATCC TCAGGAAGAT CCATTGCTTC TAGTTCTTGC 9420

GGCAGACCAC GTTATAGCTA AAGAAAGTGT TTTCTGTGAT GCTATTAAAA ATGCAACTCC 9480

CATCGCTAAT CAAGGTAAAA TTGTAACGTT TGGAATTATA CCAGAATATG CTGAAACTGG 9540

TTATGGGTAT ATTGAGAGAG GTGAACTATC TGTACCGCTT CAAGGGCATG AAAATACTGG 9600

TTTTTATTAT GTAAATAAGT TTGTCGAAAA GCCTAATCGT GAAACCGCAG AATTGTATAT 9660

GACTTCTGGT AATCACTATT GGAATAGTGG AATATTCATG TTTAAGGCAT CTGTTTATCT 9720

TGAGGAATTG AGAAAATTTA GACCTGACAT TTACAATGTT TGTGAACAGG TTGCCTCATC 9780

CTCATACATT GATCTAGATT TTATTCGATT ATCAAAAGAA CAATTTCAAG ATTGTCCTGC 9840

TGAATCTATT GATTTTGCTG TAATGGAAAA AACAGAAAAA TGTGTTGTAT GCCCTGTTGA 9900

TATTGGTTGG AGTGACGTTG GATCTTGGCA ATCGTTATGG GACATTAGTC TAAAATCGAA 9960

AACAGGAGAT GTATGTAAAG GTGATATATT AACCTATGAT ACTAAGAATA ATTATATCTA 10020

CTCTGAGTCA GCGTTGGTAG CCGCCATTGG AATTGAAGAT ATGGTTATCG TGCAAACTAA 10080

AGATGCCGTT CTTGTGTCTA AAAAGAGTGA TGTACAGCAT GTAAAAAAAA TAGTCGAAAT 10140

GCTTAAATTG CAGCAACGTA CAGAGTATAT TAGTCATCGT GAAGTTTTCC GACCATGGGG 10200

AAAATTTGAT TCGATTGACC AAGGTGAGCG ATACAAAGTC AAGAAAATTA TTGTGAAACC 10260

TGGTGAGGGG CTTTCTTTAA GGATGCATCA CCATCGTTCT GAACATTGGA TCGTGCTTTC 10320

TGGTACAGCA AAAGTAACCC TTGGCGATAA AACTAAACTA GTCACCGCAA ATGAATCGAT 10380

ATACATTCCC CTTGGCGCAG CGTATAGTCT TGAGAATCCG GGCATAATCC CTCTTAATCT 10440

TATTGAAGTC AGTTCAGGGG ATTATTTGGG AGAGGATGAT ATTATAAGAC AGAAAGAACG 10500

TTACAAACAT GAAGATTAAC ATATGAAATC TTTAACCTGC TTTAAAGCCT ATGATATTCG 10560

CGGGAAATTA GGCGAAGAAC TGAATGAAGA TATTGCCTGG CGCATTGGGC GTGCCTATGG 10620

CGAATTTCTC AAACCGAAAA CCATTGTTTT AGGCGGTGAT GTCCGCCTCA CCAGCGAAGC 10680

GTTAAAACTG GCGCTTGCGA AAGGTTTACA GGATGCGGGC GTCGATGTGC TGGATATCGG 10740

TATGTCCGGC ACCGAAGAGA TCTATTTCGC CACGTTCCAT CTCGGAGTGG ATGGCGGCAT 10800

CGAAGTTACC GCCAGCCATA ACCCGATGGA TTACAACGGC ATGAAGCTGG TGCGCGAAGG 10860

GGCTCGCCCG ATCAGCGGTG ATACCGGACT GCGCGATGTC CAGCGTCTGG CAGAAGCCAA 10920

TGACTTCCCT CCTGTCGATG AAACCAAACG TGGTCGCTAT CAGCAAATCA ATCTGCGTGA 10980

CGCTTACGTT GATCACCTGT TCGGTTATAT CAACGTCAAA AACCTCACGC CGCTCAAGCT 11040

GGTGATCAAC TCCGGGAACG GCGCAGCGGG TCCGGTGGTG GACGCCATTG AAGCCCGATT 11100

TAAAGCCCTC GGCGCACCGG TGGAATTAAT CAAAGTACAC AACACGCCGG ACGGCAATTT 11160

CCCCAACGGT ATTCCTAACC CGCTGCTGCC GGAATGCCGC GACGACACCC GTAATGCGGT 11220

CATCAAACAC GGCGCGGATA TGGGCATTGC CTTTGATGGC GATTTTGACC GCTGTTTCCT 11280

GTTTGACGAA AAAGGGCAGT TTATCGAGGG CTACTACATT GTCGGCCTGC TGGCAGAAGC 11340

GTTCCTCGAA AAAAATCCCG GCGCGAAGAT CATCCACGAT CCACGTCTCT CCTGGAACAC 11400
CGTTGATGTG GTGACTGCCG CAGGCGGCAC CCCGGTAATG TCGAAAACCG GACACGCCTT 11460
TATTAAAGAA CGTATGCGCA AGGAAGACGC CATCTACGGT GGCGAAATGA GCGCTCACCA 11520
TTACTTCCGT GATTTCGCTT ACTGCGACAG CGGCATGATC CCGTGGCTGC TGGTCGCCGA 11580
ACTGGTGTGC CTGAAAGGAA AAACGCTGGG CGAAATGGTG CGCGACCGGA TGGCGGCGTT 11640
TCCGGCAAGC GGTGAGATCA ACAGCAAACT GGCGCAACCC GTTGAGGCAA TTAATCGCGT 11700
GGAACAGCAT TTTAGCCGCG AGGCGCTGGC GGTGGATCGC ACCGATGGCA TCAGCATGAC 11760
CTTTGCCGAC TGGCGCTTTA ACCTGCGCTC CTCCAACACC GAACCGGTGG TGCGGTTGAA 11820
TGTGGAATCA CGCGGTGATG TAAAGCTAAT GGAAAAGAAA ACTAAAGCTC TTCTTAAATT 11880
GCTAAGTGAG TGATTATTTA CATTAATCAT TAAGCGTATT TAAGATTATA TTAAAGTAAT 11940
GTTATTGCGG TATATGATGA ATATGTGGGC TTTTTTATGT ATAACGACTA TACCGCAACT 12000
TTATCTAGGA AAAGATTAAT AGAAATAAAG TTTTGTACTG ACCAATTTGC ATTTCACGTC 12060
ACGATTGAGA CGTTCCTTTG CTTAAGACAT TTTTTCATCG CTTATGTAAT AACAAATGTG 12120
CCTTATATAA AAAGGAGAAC AAAATGGAAC TTAAAATAAT TGAGACAATA GATTTTTATT 12180
ATCCCTGTTT ACGATATTAT AGCCAAAGTT GTATCCTGCA TCAGTCCTGC AATATTTCAC 12240
GAGTGCTTTG TTAACTGAAT ACATGTCTGC CATTTTCCAG ATGATAACGA CGTCATCGCA 12300
ATTGATGGTA AAACACTTCG GCACACTTAT GACAAGAGTC GTCGCAGAGG AGTGGTTCAT 12360
GTCATTAGTG CGTTTCAGCA ATGCACAGTC TGGTCCTCGG ATAGATCAAG ACGGATGAGA 12420
AACCTAATGC GTTCACAGTT ATTCATGAAC TTTCTAAAAT GATGGGTATT AAAGGAAAAA 12480
TAATCATAAC TGATGCGATG GCTTGCCAGA AAGATATTGC AGAGAAGATA TAAAAACAGA 12540
GATGTGATTA TTTATTCGCT GTAAAAGGAA ATAAGAGTCG GCTTAATAGA GTCTTTGAGG 12600
AGATATTTAC GCTGAAAGAA TTAAATAATC CAAAACATGA CAGTTACGCA ATTAGTGAAA 12660
AGAGGCACGG CAGAGACGAT GTCCGTCTTC ATATTGTTTG AGATGCTCCT GATGAGCTTA 12720
TTGATTTCAC GTTTGAATGG AAAGGGCTGC AGAATTTATG AATGGCAGTC CACTTTCTCT 12780
CAATAATAGC AGAGCAAAAG AAAGAATCCG AAATGACGAT CAAATATTAT ATTAGATCTG 12840
CTGCTTTAAC CGCAGAGAAG TTCGCCACAG TAAATCGAAA TCACTGGCGC ATGGAGAATA 12900
AGTTGCACAG TAGCCTGATG TGGTAATGAA TGAAATCGAC TATAATATAA GAAGGCGAGT 12960
TGCATTCGAA TGATTTTCTA GAATGCGGCA CATCGCTATT AATATCTGAC AATGATAATG 13020
TATTCAAGGC AGGATTATCA TGTAAGATGC GAAAAGCAGT CATGGACAGA AACTTCCTAG 13080
CGTCAGGCAT TGCAGCGTGC GGGCTTTCAT AATCTTGCAT TGGTTTTGAT AAGATATTTC 13140
TTTGGAGATG GGAAAATGAA TTTGTATGGT ATTTTTGGTG CTGGAAGTTA TGGTAGAGAA 13200
ACAATACCCA TTCTAAATCA ACAAATAAAG CAAGAATGTG GTTCTGACTA TGCTCTGGTT 13260
TTTGTGGATG ATGTTTTGGC AGGAAAGAAA GTTAATGGTT TTGAAGTGCT TTCAACCAAC 13320
TGCTTTCTAA AAGCCCCTTA TTTAAAAAAG TATTTTAATG TTGCTATTGC TAATGATAAG 13380

ATACGACAGA GAGTGTCTGA GTCAATATTA TTACACGGGG TTGAACCAAT AACTATAAAA 13440

CATCCAAATA GCGTTGTTTA TGATCATACT ATGATAGGTA GTGGCGCTAT TATTTCTCCC 13500

TTTGTTACAA TATCTACTAA TACTCATATA GGGAGGTTTT TTCATGCAAA CATATACTCA 13560

TACGTTGCAC ATGATTGTCA AATAGGAGAC TATGTTACAT TTGCTCCTGG GGCTAAATGT 13620

AATGGATATG TTGTTATTGA AGACAATGCA TATATAGGCT CGGGTGCAGT AATTAAGCAG 13680

GGTGTTCCTA ATCGCCCACT TATTATTGGC GCGGGAGCCA TTATAGGTAT GGGGGCTGTT 13740

GTCACTAAAA GTGTTCCTGC CGGTATAACT GTGTGCGGAA ATCCAGCAAG AGAAATGAAA 13800

AGATCGCCAA CATCTATTTA ATGGGAATGC GAAAACACGT TCCAAATGGG ACTAATGTTT 13860

AAAATATATA TAATTTCGCT AATTTACTAA ATTATGGCTT CTTTTTAAGC TATCCTTTAC 13920

TTAGTTATTA CTGATACAGC ATGAAATTTA TAATACTCTG ATACATTTTT ATACGTTATT 13980

CAAGCCGCAT ATCTAGCGGT AACCCCTGAC AGGAGTAAAC AATG 14024

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12441 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Salmonella enterica serovar muenchen serogroup C2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GTTGACAAAT ACCGACCGTA TAATGAATCA AACGTTCTGG ATTGGTATTT ATCCAGGCTT 60

GACTACAGAG CATTTAGATT ATGTCGTAAG TAAGTTTGAA GAATTTTTTG GTTTAAATTT 120

CTAATTTTTA GGATAGGATG CTTGATGTGA ATAAGAAAAT CCTAATGACT GGCGCTACTA 180

GCTTTGTAGG TACCCATCTA CTACATAGTC TCATAAAGGA AGGTTATAGT ATTATTGCAT 240

TAAAGCGTCC TATAACCGAG CCAACGATTA TCAATACCTT GATTGAATGG TTGAATATAC 300

AAGATATAGA AAAAATATGT CAATCATCTA TGAATATTCA TGCGATTGTC CATATTGCAA 360

CAGACTATGG TCGAAACAGA ACCCCTATAT CTGAACAATA TAAATGTAAT GTCCTATTAC 420

CAACAAGACT GCTTGAGTTA ATGCCAGCGC TTAAAACGAA ATTCTTTATT TCTACTGACT 480

CTTTTTTTGG GAAATATGAG AAGCACTATG GATATATGCG TTCTTACATG GCATCTAAAA 540

GACATTTTGT AGAACTATCA AAAATATACG TAGAGGAACA TCCAGACGTT TGTTTTATAA 600

ATTTACGTTT AGAACATGTT TACGGTGAGA GGGATAAAGC AGGTAAAATA ATCCCGTATG 660

TTATCAAAAA AATGAAAAAC AATGAAGATA TTGATTGTAC GATCGCCAGG CAGAAAAGAG 720

ATTTTATTTA TATAGACGAT GTTGTTTCGG CCTATTTGAA AATTTTAAAG GAGGGTTTTA 780

ACGCTGGACA CTATGATGTC GAGGTGGGGA CTGGAAAATC GATAGAGCTA AAAGAAGTGT 840
TTGAGATAAT AAAAAAAGAA ACGCATAGTA GTAGTAAGAT AAATTATGGT GCAGTTGCGA 900
TGCGTGATGA TGAGATTATG GAGTCACATG CAAATACCTC TTTCTTGACT CGATTAGGTT 960
GGAGTGCCGA GTTTTCTATT GAGAAGGGTG TGAAAAAAAT GTTGAGTATG AAAGAGTAAT 1020
GAATCGTATT ATTAGAATGT TAGGTGTAGA TAAAGCAATT CGTTATGTTA TTTTTGGTAA 1080
GATAATATCT GTATTAACGG GTTTACTGTT AATAATGTTA ATATCACACC ATTTATCTAA 1140
AGACGCACAG GGCTATTATT ATACATTTAA TTCAGTAGTG GCACTACAGA TAATATTTGA 1200
ATTGGGGCTA TCAACGGTAA TCATTCAATT CGCTAGCCAT GAAATGTCAG CGTTAAAATA 1260
TGATTATTCT GAACGAGATA TTATAGGTGA AAGTAAAAAT AAGCAACGTT ACCTATCGTT 1320
ATTTCGGTTG GCAATAAAAT GGTATGCAGT AATAGCTTTG CTAATAATAT TAATAGTCGG 1380
TCCCATCGGG TATGTTTTTT TTACGCAAAA AGAAGGCTTA GGTGTACCTT GGCAAGGGGC 1440
ATGGTTATTA TTAACAATAG TTACAGCTTT TAATATTTTT CTTGTTTCTG TACTTTCTGT 1500
CGCTGAAGGG AGTGGGTTAA TTACTGATGT GAATAAAATG AGAATGTATC AGTCGCTGTT 1560
AGCTGGTATA TTGGCAGTAA GCTTACTTAT TAGTGGCTTT GGACTATATG CTACGTCTGC 1620
AATAGCTATT TCAGGGACTA TCATATTCTC CATATTTTCA TATAAGTATT TTAAAAAAAT 1680
TTTCCTGCAA TCTTTAAAGC ATAAAAATAA ATATACTGAA GGTGGTATTT CATGGGTTAA 1740
TGAAATATTT CCTATGCAAT GGCGAATTGC TCTAAGTTGG ATGTCAGGGT ATTTTATTTA 1800
TTTTGTTATG ACCCCCATTG CATTCAAATA TTTCGGGGCT ATATATGCAG GGCAGTTAGG 1860
GATGTCTTTA ACATTATGCA ATATGGTAAT GGCTACGGGC CTGGCTTGGA TATCCACTAA 1920
ATATCCAAAA TGGGGAGTAA TGGTTTCCAA CAAACAGCTT GCGGAACTGA GTAAATCGTT 1980
CAAAAGTGCA GTAATGCAAT CATCCTTTTT TGTCTTGACA GGATTAACTG GTGTATACAT 2040
TTCATTATGG TTATTGAAAT TATCTGGTTC AAACATTGGC GAGCGGTTTT TGGGATTGCA 2100
GGATTTTTTC TTTTTATCTT TAGCAATTAT TGGTAATCAC ATTGTAGCTT GCTTTGCAAC 2160
CTATATAAGA GCGCATAAAA CTGAAAAAAT GACATTGGCA TCATGTATAA TGGCTCTCTT 2220
GACTATAACT ACAATGTTGT TTGTTGCATA TTTAGAGTAC TCGAGGTTCT ACATGTTAAT 2280
GTATGCAGCA CTAACGTGGT TATATTTTGT TCCTCAAACT TATATAATCT TTAAAAGATT 2340
CAAGAGTTCT TATGAGTAAA AAACCTCTTC TTACTATTGC TATTCCGACA TATAACCGCT 2400
CTTCATGTTT GGCTCGTTTA CTTGATAGTA TAATTCAACA GGAGAACTAT TGTCATGATG 2460
AACTCGAGGT TATTGTTTGT GATAATGCTT CAACAGATGA AACAGCAAGA ATAGCCAAGA 2520
GTGGCTTAGA TAAAATAAGA AATAGTACTT ATCATCTAAA TGAAGAAAAC TTAGGAATGG 2580
ATGGTAACTT CCAGAAATGT TTTGAGTTAT CAAATGGAAA ATATCTTTGG ATGATTGGCG 2640
ATGATGATCT AATAGTCAAA AATGGTATTT CGAAGGTTTT TTCGATATTA AAGTCCCGGC 2700
CTGCATTAGA TATGGTGTAT GTAAATTCAG CAGCAAAGAC TGAGTTAAAC TATAATGCTG 2760
ATGTGAGGAC GTCATTCTAC ACAAATGATG TAGATTTTAT TTCAGACGTG AAAGTTATGT 2820

TCACGTTTAT TTCTGGAATG ATATGTAAGA AAACTGATGC AATTGTCAAA GCCGTTGGTA 2880

TTTTCAGTCC GCAAACTACT GGAAAATATC TTATGCATTT AACATGGCAA TTGCCATTAC 2940

TTAAACAGGG TGGAGAGTTC GCAGTTATCC ATAATAATAT AATTGAGGCT GAGCCAGATA 3000

ATTCAGGTGG ATATCATTTA TATAAGGTTT TTTCTAATAA TCTTGCGACA ATCTTTGATG 3060

TTTTTTATCC CAGAGAGCAC CGTGTAAGTA AAAGAGTTCG CGCATCAGCA TGTTTATTCT 3120

TACTTAACTT CATAGGCGAT GAAGATAAAA CCAAAAATTT TGCTACAAAT AATTATTTAA 3180

GAGATTGCGA TAGTGCATTT ATAGATTTAA TTATATATAA ATATGGGCTT AGGTTTTTCT 3240

ATCTATATCC TAAAACTGTG CCTTTATTTA GAAAAATAAA ATATATTATA AAGACGGTTT 3300

TAATGCGGAA ATAAAAATTA TTCAAGATGG TTTGCTGAAA ACGACTTATA GGACTATCTA 3360

ATGTTTGTCT ATAGTTTAAG ATTAAAATTA AATCTTATCA TATCATTATT GAGTAAAGTT 3420

AGGCGGAAAT CAAAAGCAAA GTTTCTTGTT CTGCTTAGCG GATATGATTT TAAAATGGTT 3480

GGGAAGAATT TTAAATTGAA TGTCAAACCT TACTCTGCAA AAAATAACAC CTCTTCCAAA 3540

TGGGGTAGTA TGCGGGTTGG TGATAACTGC TGGATTGAAG CTGTATATAA TTATGGTGAT 3600

GAAAAATTTG AACCTTATTT GTACATAGGT GATCGTATAT GTTTAAGTGA TAATGTTCAT 3660

ATTTCTTGCG TATCATGTTT AATTTTAGAA AACGATATAT TAATTGGTAG CAAAGTTTAT 3720

ATAGGCGATC ATAGCCATGG CAGTTATAAA GTATGCAGTC CGAAAATAGA ACCGCCAGCA 3780

AATAAGCCAT TAGGTGATAT TGCTCCTATT AAAATAGGTA ATTGCTGCTG GATTGGAGAT 3840

AATGCAGTAA TTCTGGCTGG TAGTGAAATT TGTGATGGCT GTGTAATCGC AGCTAATTCA 3900

GTCGTCAAGG ATTTAAAAGT CGATAAGCCA TGTTTAATTG GTGGGGTTCC TGCTAAAGTA 3960

ATAAAGGTAT TTTAAAATGA ATGTTTTTAT CAGTATTTGT ATACCGTCTT ATAATAGAGC 4020

TGAGTTTTTA GAGCCACTAC TGGATAGCAT ATATAATCAA GATTATTGTT TAAAGAATAA 4080

TGATTTTGAG GTCATTGTTT GTGAAGATAA ATCTCCACAG AGAGATGAGA TAAACTCTAT 4140

TATCGAAAAC TATAAAGCAA AAAATAATAA ACAAAATCTT TATGTTAATT TCAATGAAGA 4200

TAATTTAGGC TATGATAAGA ATTTAAAAAA ATGCATTAGT TTGACGACAG GTAAATATTG 4260

CATGATCATG GGCAACGATG ATCTATTAGC AGATGGAGCG TTATCAAAAA TAGTGAAAGT 4320

TTTGAAGGCT AATCCTGAAA TTGTATTGGC TACGCGAGCG TATGGTTGGT TTAAGGAAAA 4380

TCCGAATGAG TTATGTGATA CTGTTCGTCA TTTAACAGAC GATACTTTAT TTCAGCCGGG 4440

GGCTGATGCC ATTAAATTTT TCTTCCGTAG AGTTGGAGTT ATTTCAGGCT TTATTGTCAA 4500

TGCTGAAAAA GCAAAAAAAC TATCGAGTGA TTTATTTGAT GGGCGTTTAT ATTATCAAAT 4560

GTACCTTGCT GGTATGCTAA TGGCTGAAGG TCAGGGATAC TATTTTAGCG ACGTGATGAC 4620

ATTGTCGAGG GATACAGAGG CTCCTGACTT TGGTAACGCT GGAACTGAAA AAGGAGTTTT 4680

CACCCCGGGG GGGTATAAAC CAGAGGGCCG TATACATATG GTTGAAGGCT TGTTGCTAAT 4740

TGCAAAATAT ATAGAAGATA CAACAAAAAT TGATGGCGTT TATGCTGGAA TTAGAAAAGA 4800

CTTAGCGAAC TATTTTTATC CTTATATTCG AGATCAACTC GACTTGCCTC TTTATACTTA 4860

TATTAAAATG ATAAATAAAT TTCGGAAAAT GGGATTTTCA AATGAAAAGC TTTTCTATGT 4920

GCATGCCTTT TTAGGGTATG TACTAAAACG GAGGGGCTAT GATGCTTTAA TTAAATACAT 4980

TCGTAGCAAA AAAGGCGGTA CTCCGCGTCT TGGTATTTAA CCTCCACTTT CAAAAAATGT 5040

TATGAATATA CTTCTTGCTG CGATATTAGG CGTTAACTTA TTTTCTCCAT ATATTAGTTC 5100

GTGGATGGTG GGTATGCTGC CATTTCCACC AGGAGCAATC CTAAGGGATG TACTCAATGT 5160

ATTTTTTGTG GCGTTAGTGC TAGTTCGATT TGTCATTGAT AGGAAAAAAA CTTATTTCCC 5220

GTTGGTTTTT ACTATTTTTT CATGGTCGGC GGTAATACTA TGGGTAATAG CGTTAACTAT 5280

ATTCTCACCG GATAAAATTC AAGCAATTAT GGGGGGGCGG AGTTATATTT TATTCCCGGC 5340

AGTTTTCATA GCATTAGTGA TTTTAAAAGT ATCATACCCG CAATCCTTAA ATATTGAAAA 5400

AATAGTTTGC TACATAATTT TTCTAATGTT TATGGTTGCG ACAATATCTA TTATTGATGT 5460

ACTAATGAAT GGAGAGTTCA TTAAATTGCT CGGATATGAT GAGCATTATG CAGGAGAACA 5520

ATTAAACTTA ATTAATAGCT ATGATGGGAT GGTCCGGGCT ACAGGCGGTT TTAGTGATGC 5580

TCTCAATTTT GGATATATGC TCACATTAGG TGTTTTGTTA TGTATGGAGT GTTTTTCCCA 5640

AGGATATAAA AGATTATTGA TGCTTATTAT TAGTTTTGTG CTATTTATAG CGATCTGCAT 5700

GAGTCTTACT AGAGGAGCAA TACTTGTTGC TGCGCTTATT TACGCACTTT ATATAATTTC 5760

AAATCGGAAG ATGCTTTTTT GTGGAATAAC TTTATTTGTA ATAATTATAC CCGTTTTAGC 5820

AATTTCTACT AATATTTTTG ACAACTATAC AGAAATTTTG ATCGGCAGGT TTACAGATTC 5880

GTCTCAGGCA TCGCGTGGAT CTACACAGGG GCGGATAGAT ATGGCAATTA ATTCATTAAA 5940

CTTCCTGTCA GAACATCCAT CAGGTATAGG TCTGGGTACT CAAGGTTCAG GAAACATGCT 6000

TTCGGTAAAA GATAATAGGT TAAATACGGA TAATTATTTT TTCTGGATCG CCCTTGAGAC 6060

TGGTATTATT GGCTTAATCA TAAATATTAT TTATCTGGCA AGTCAATTTT ATTCTTCAAC 6120

TTTACTAAAT AGAATATATG GCAGTCATTG TAGCAATATG CACTATAGAT TATATTTTCT 6180

CTTTGGAAGT ATATATTTTA TAAGTGCAGC GTTAAGTTCA GCACCTTCGT CATCAACTTT 6240

TTCTATATAT TATTGGACAG TTTTAGCTTT GATTCCATTT TTAAAATTAA CAAATAGACG 6300

GTGCACGCGA TAATGAATAA TAAAAAGGTT TTGATGGATA TTAGTTGGTC TAATAAAGGG 6360

GGGATTGGAC GTTTTACTGA TGAAATTTCT AAACTACTAT GTGATATATC TAAGGAGGAA 6420

CTATATAGAA AATGTGCTTC TCCGCTGGCC CCATTAGGTT TAGCAGTCAA TATTTTTCTG 6480

CGAAAGAAAA CTGATGTGGT TTTTCTTCCT GGCTATATTC CACCACTTTT TTGTTCGAAA 6540

AAGTTCATAA TAACAATACA TGATCTAAAT CATCTGGATT TAAATGATAA TTCCTCTCTT 6600

TTTAAGAGGT TATTTTATAA TTTTATAATA AAGCGCGGTT GTAGAAAAGC ATATAAAATA 6660

TTTACAGTTT CGAATTTTTC AAAAGAAAGA ATAGTAGCAT GGTCAGGTGT AAACCCTAAT 6720

AAAATAGTCA CGGTATATAA TGGGGTATCT AGTCTATTTA ATGCCGATGT AAAACCATTG 6780

AATTTAGGCT ATAAATATTT GCTATGTGTA GGAAACAGAA AAACTCATAA GAATGAGAAG 6840

TGTGTTATAT CTGCCTTTGC CAAAGCAGAT ATTGATCCAT CAATAAAACT CGTTTTTACT 6900

GGTAATCCTT GTAATGATTT AGAAAAACTA ATAATACAAC ATGGTTTAAG TGAACGTGTA 6960

AAGTTCTTTG GGTTCGTGTC TGAAAAAGAT TTACCATCGT TATATAAGGG CTCGTTAGGA 7020

TTAGTTTTCC CTTCTTTATA TGAAGGTTTT GGATTACCTG TAGTGGAGGG CATGGCCTGT 7080

GGTATTCCTG TATTAACTTC TCTAACTTCA TCATTGCCAG AGGTGGCTGG AGATGCAGCG 7140

ATTCTTGTCG ACCCTCTTTC GGAAGATGCT ATTACTAAAG GAATTTCGAG GTTAATTAAT 7200

GATTCTGAAC TTCGTAAGCA TTTAATCCAA AAGGGGCTTT TGCGGGCAAA GAGGTTCAAT 7260

TGGCAAAACG TGGTTAGTGA GATTGAAATG GTACTGACAG AGGCATGTGA TGGAAATAAA 7320

TGAAATAAAA ATATCTCTCG TTCATGAGTG GTTATTAAGT TATGCAGGCT CCGAACAGGT 7380

ATCATCTGCC ATCCTGCATG TTTTTCCTGA AGCGAAGTTA TATTCGGTGG TTGATTTTCT 7440

AACGGATGAA CAAAGAAGAC ATTTTCTGGG GAAATATGCG ACTACCACAT TTATTCAAAA 7500

TTTACCTAAA GCTAAAAAAT TTTACCAGAA ATATTTACCA CTAATGCCAC TGGCTATTGA 7560

ACAACTTGAT TTATCAGATG CTAATATCAT CATTAGTAGC GCCCATTCCG TTGCAAAAGG 7620

TGTTATTTCC GGACCAGATC AGCTTCACAT TAGCTATGTT CATTCTCCTA TTCGATATGC 7680

GTGGGATTTA CAGCATCAGT ACCTTAATGA GTCTAACCTG AATAAAGGAA TTAAAGGTTG 7740

GTTAGCAAAA TGGCTTCTTC ACAAAATACG AATTTGGGAT TCTCGAACCG CAAATGGGGT 7800

TGATCATTTT ATAGCTAATT CTCAATATAT CGCGCGTAGA ATTAAAAAAG TATACAGACG 7860

TGAGGCTTCA GTTATATATC CGCCTGTAGA TGTGGATAAT TTTGAAGTAA AAAATGAAAA 7920

GCAAGACTAT TATTTCACAG CATCCCGTAT GGTACCCTAC AAACGTATTG ATCTTATTGT 7980

CGAAGCCTTT AGTAAAATGC CGGAAAAGAA ATTAGTAGTT ATTGGTGATG GACCGGAGAT 8040

GAAAAAAATA AAGAGCAAGG CTACAGACAA TATAAAATTG CTCGGTTATC AATCTTTTCC 8100

TGTTTTAAAA GAGTATATGC AGAGCGCCAG GGCGTTTGTT TTTGCAGCGG AAGAGGACTT 8160

TGGAATAATA CCTGTCGAAG CTCAAGCTTG CGGTACCCCT GTTATTGCCT TTGGGAAGGG 8220

TGGGGCCTTA GAAACCGTTC GCCCACTAGG TGTAGAGGAA CCGACTGGCA TTTTCTTCAA 8280

GGAACAGAAT ATTGCTTCTT TGCATGAAGC TGTTAGTGAA TTTGAAAAAA ATGCATCATT 8340

TTTTACATCT CAGGCTTGTA GAAAAAATGC AGAAAAATTT TCTCGATCAA GATTTGAACA 8400

AGAATTTAAG AACTTTGTTA ATGAAAAGTG GAATCTTTTC AAAACAGAAC AGATTATTAA 8460

ACGTTAATTA TGGTTTATTG AATGTCTAAA TTAATACCAG TAATAATGGC CGGTGGGATT 8520

GGTAGCCGTT TGTGGCCACT TTCACGTGAA GAGCATCCGA AACAGTTTTT AAGCGTAGAT 8580

GGTGAATTAT CTATGCTGCA AAACACCATT AAAAGATTGA CTCCTCTTTT GGCTGGAGAA 8640

CCTTTAGTCA TTTGTAATGA TAGTCACCGC TTCCTTGTCG CTGAACAACT TCGAGCTATA 8700

AATAAACTAG CAAATAACAT CATATTAGAG CCAGTGGGGC GTAATACAGC CCCAGCTATA 8760

GCGCTGGCCG CTTTTTGTTC ACTTCAGAAT GTCGTCGATG AAGACCCGCT TTTGCTTGTC 8820

CTTGCTGCGG ATCATGTCAT CCGCGATGAG AAAGTGTTTC TTAAAGCTAT CAATCACGCT 8880

GAATTTTTTG CAACACAAGG TAAGCTAGTA ACGTTTGGTA TTGTACCCAC ACAGGCCGAA 8940

ACTGGCTACG GTTATATTTG TAGAGGTGAA GCAATCGGGG AAGATGCTTT TTGTGTAGCC 9000
GAATTTGTAG AGAAGCCTGA TTTCGATACA GCGCGTCATT ATGTAGAATC AGAGAAATAT 9060
TATTGGAACA GCGGTATGTT CCTATTTCGT GCAAGTAGTT ACTTACAAGA ATTAAAGGAT 9120
CTGTCCCCCG ATATTTACCA AGCATGTGAA AATGCGGTAG GGAGTATTAA TCCTGATCTT 9180
GATTTTATCC GTATTGATAA AGAAGCATTC GCAATGTGCC CTAGTGATTC TATCGATTAT 9240
GCGGTAATGG AACATACTAG GCATGCAGTT GTCGTACCGA TGAATGCCGG CTGGTCAGAT 9300
GTGGGGTCAT GGTCTTCACT GTGGGATATT TCTAAGAAAG ATCCACAACG TAATGTATTA 9360
CATGGCGATA TTTTTGCATA TAATAGTAAA GATAATTATA TCTATTCTGA AAAATCGTTT 9420
ATTAGTACAA TCGGAGTAAA TAATTTAGTT ATCGTGCAGA CAGCAGATGC ATTATTAGTA 9480
TCTGATAAAG ATTCAGTCCA GGATGTTAAA AAAGTTGTTG ATTATTTAAA AGCTAATAAT 9540
AGAAACGAAC ATAAAAAACA TTTAGAGGTT TTCCGACCGT GGGGAAAATT TAGCGTAATT 9600
CATAGTGGCG ATAATTATTT AGTTAAAAGA ATAACTGTTA AACCAGGCGC GAAGTTTGCT 9660
GCTCAGATGC ATCTCCATCG TGCTGAGCAT TGGATAGTGG TATCTGGTAC TGCTTGTATT 9720
ACTAAGGGGG AAGAAATTTT TACAATTTCG GAGAATGAAT CAACATTTAT ACCTGCTAAT 9780
ACAGTTCATA CGTTAAAAAA CCCCGCGACT ATTCCATTAG AACTAATAGA AATTCAATCT 9840
GGCACCTATC TTGCGGAGGA TGATATTATT CGCCTGGAGA AACATTCTGG ATATCTGGAG 9900
TAATGAATTG ATGAAAAATA TATATAATAC TTACGATGTT ATCAACAAAT CTGGAATTAA 9960
TTTTGGAACC AGTGGTGCCC GCGGCCTTGT TACCGATTTT ACACCCGAAG TTTGCGCACG 10020
ATTTACCATT TCCTTTTTGA CAGTAATGCA GCAAAGATTC TCATTTACAA CGGTTGCGCT 10080
CGCAATTGAT AATCGTCCAA GCAGTTACGC GATGGCTCAA GCTTGTGCCG CTGCTTTGCA 10140
AGAAAAAGGA ATTAAAACCG TTTACTATGG CGTAATTCCA ACACCTGCTT TAGCTCATCA 10200
ATCAATTTCC GATAAAGTAC CTGCAATCAT GGTTACTGGC AGTCATATCC CTTTTGACCG 10260
TAATGGCCTG AAATTTTATA GACCAGATGG TGAAATTACT AAAGATGATG AGAATGCTAT 10320
TATTCATGTT GATGCCTCAT TTATGCAGCC TAAGCTTGAA CAATTGACAA TTTCCACAAT 10380
CGCTGCTAGA AATTATATTC TACGATATAC CTCATTATTT CCAATGCCAT TCTTGAAAAA 10440
TAAGCGCATT GGAATTTATG AGCATTCTAG TGCGGGTCGT GATCTCTATA AGACGTTATT 10500
CAAAATGTTG GGTGCTACAG TTGTTAGTTT AGCAAGGAGC GACGAATTTG TTCCTATTGA 10560
TACTGAAGCT GTAAGTGAAG ATGATAGAAA TAAAGCAATC ACATGGGCAA AAAAATATCA 10620
GTTAGATGCT ATATTTTCAA CTGATGGTGA TGGAGATCGC CCTCTGATAG CTGACGAATA 10680
TGGAAATTGG TTAAGAGGAG ATATATTAGG CCTTCTGTGC TCTCTCGAAT TAGCTGCTGA 10740
TGCAGTCGCT ATTCCTGTAA GCTGCAACAG TACAATCTCA TCTGGTAACT TTTTTAAACA 10800
TGTGGAACGA ACAAAGATTG GTTCACCCTA TGTGATTGCA GCATTTGCTA AATTATCTGC 10860
AAACTATAAT TGTATAGCTG GTTTTGAAGC GAATGGTGGC TTTCTGCTAG GTAGCGATGT 10920
TTATATTAAT CAGCGTTTAC TTAAGGCATT ACCAACACGT GATGCTTTAT TACCTGCCAT 10980

TATGCTTCTG TTTGGTAGCA AGGACAAAAG TATTAGTGAG CTTGTTAAAA AACTTCCTGC 11040

TCGCTATACC TATTCAAACA GATTACAGGA TATAAGTGTT AAAACAAGTA TGTCTTTAAT 11100

AAATCTTGGT CTGACAGATC AAGAGGATTT TTTGCAGTAT ATTGGTTTTA ATAAACATCA 11160

TATATTACAT TCTGATGTTA CTGATGGCTT TAGAATCACT ATCGATAACA ACAATATTAT 11220

TCATTTACGA CCTTCAGGCA ATGCCCCTGA GTTGCGTTGC TATGCGGAGG CTGACTCGCA 11280

AGAGGATGCA TGTAATATTG TTGAAACTGT TCTCTCTAAT ATCAAAAGCA AACTGGGTAG 11340

AGCTTAATGC TGTTGATAAT AGAGCGTTTC TTTCCAGTAA TACTTTGTCT GGTTATCTGG 11400

TACCCAAGTT GAGGGTGAGA ATTAAATGGA TCGTTTTGAT AATAAGTATA ACCCAAATTT 11460

ATGCAAAATA TTATTGGCTA TATCAGATTT ACTGTTTTTT AATGTAGCCT TATGGGCATC 11520

GTTAGGAGTT GTATATTTAA TCTTTGATGA AGTTCAGCGA TTTGTACCAC AAGAGCAATT 11580

AGATAATCGA TTTATATCAC ATTTTATTCT ATCTATAGTA TGCGTTGGAT GGTTTTGGGT 11640

TCGACTGCGT CACTATACAT ATCGAAAGCC ATTCTGGTAT GAGTTGAAAG AGGTTATTCG 11700

TACTATCGTT ATTTTTGCTG TGTTTGATTT GGCTTTAATT GCGTTTACAA AATGGCAGTT 11760

TTCACGCTAT GTCTGGGTGT TTTGTTGGAC TTTTGCCATA ATCCTGGTGC CTTTTTTTCG 11820

CGCACTTACA AAGCATTTAT TGAACAAGCT AGGTATCTGG AAGAAAAAAA CTATCATCCT 11880

TGGGAGCGGA CAGAATGCTC GTGGTGCATA TTCTGCGCTG CAAAGTGAGG AGATGATGGG 11940

GTTTGATGTT ATCGCTTTTT TTGATACGGA TGCGTCAGAT GCTGAAATAA ATATGTTGCC 12000

GGTGATAAAG GACACTGAGA CTATTTGGGA TTTAAATCGT ACAGGTGATG TCCATTATAT 12060

CCTTGCTTAT GAATACACCG AGTTGGAGAA AACACATTTT TGGCTACGTG AACTTTCAAA 12120

ACATCATTGT CGTTCTGTTA CTGTCGTCCC CTCGTTTAGA GGATTGCCAT TATATAATAC 12180

TGATATGTCT TTTATCTTTA GCCATGAAGT TATGTTATTA AGGATACAAA ATAACTTGGC 12240

TAAAAGGTCG TCCCGTTTTC TCAAACGGAC ATTTGATATT GTTTGTTCAA TAATGATTCT 12300

TATAATTGCA TCACCACTTA TGATTTATCT GTGGTATAAA GTTACTCGAG ATGGTGGTCC 12360

GGCTATTTAT GGTCACCAGC GAGTAGGTCG GCATGGAAAA CTTTTTCCAT GCTACAAATT 12420

TCGTTCTATG GTTATGAATT C 12441

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22080 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: S. enterica serovar typhimurium (serogroup B)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAATTCGGGA GGCGCAATGA AAGTCAGCTT TTTTCTGCTG AAATTTCCAC TCTCATCGGA 60

AACCTTTGTG CTGAATCAGA TTACTGCGTT TATTGATATG GGCCATGAGG TGGAGATTGT 120

CGCGTTACAA AAAGGCGATA CCCAACATAC TCACGCCGCC TGGGAGAAGT ATGGCCTGGC 180

GGCGAAAACC CGCTGGTTAC AGGATGAGCC CCAGGGACGG CTGGCGAAAC TGCGCTACCG 240

GGCATGTAAA ACGCTGCCGG GGCTGCATCG GGCGGCGACC TGGAAAGCGC TCAATTTTAC 300

CCGCTATGGC GATGAATCAC GCAATTTGAT CCTTTCCGCG ATTTGCGCGC AGGTGAGCCA 360

GCCTTTTGTG GCGGATGTGT TTATCGCACA CTTTGGTCCG GCGGGCGTGA CGGCGGCCAA 420

ACTACGCGAA CTGGGCGTGC TTCGCGGCAA AATCGCGACT ATTTTCCACG GGATTGATAT 480

CTCTAGTCGT GAGGTGCTCA GTCATTACAC GCCGGAGTAT CAGCAGTTGT TTCGTCGTGG 540

CGATCTGATG CTGCCCATCA GCGATCTGTG GGCCGGTCGC CTGAAAAGTA TGGGCTGTCC 600

GCCGGAAAAG ATTGCCGTTT CGCGCATGGG CGTCGACATG ACGCGTTTTA CCCATCGTTC 660

GGTGAAAGCG CCAGGGATGC CGCTGGAGAT GATTTCCGTC GCGCGCCTGA CAGAAAAAAA 720

AGGCCTGCAT GTGGCGATTG AAGCCTGTCG GCAACTGAAA GCACAGGGCG TGGCGTTTCG 780

CTACCGCATT CTGGGGATTG GCCCGTGGGA ACGTCGGCTG CGCACGCTCA TCGAGCAGTA 840

TCAGCTAGAG GATGTCATTG AGATGCCGGG GTTTAAACCG AGCCATGAAG TGAAGGCGAT 900

GCTGGATGAC GCCGATGTTT TTTTGCTGCC GTCGATTACC GGTACGGATG GCGATATGGA 960

AGGTATTCCG GTAGCGCTGA TGGAGGCGAT GGCGGTAGGG ATTCCCGTGG TATCTACCGT 1020

GCATAGCGGT ATTCCGGAAC TGGTGGAGGC CGGCAAATCC GGCTGGCTGG TGCCGGAAAA 1080

CGATGCGCAG GCGCTGGCGG CCCGACTCGC TGAGTTCAGC CGGATTGACC ACGACACGCT 1140

GGAGTCGGTG ATCACGCGCG CCCGTGAAAA AGTGGCGCAA GATTTTAATC AGCAGGCGAT 1200

TAATCGCCAG TTAGCCAGCC TGCTACAAAC GATATAAACG AGGTGGTATG CCCGCGACTA 1260

AATTCTCCCG ACGTACCCTC CTGACGGCAG GTTCTGCGCT TGCTGTTCTT CCTTTTCTGC 1320

GCGCCTTGCC GGTACAGGCG CGTGAACCTC GCGAGACCGT CGATATTAAG GATTATCCGG 1380

CGGATGACGG TATCGCCTCG TTCAAACAGG CCTTCGCCGA CGGACAGACC GTGGTCGTAC 1440

CGCCAGGATG GGTGTGTGAA AATATCAATG CGGCGATAAC GATTCCGGCG GGAAAAACGC 1500

TGCGGGTACA GGGCGCGGTG CGTGGGAATG GCCGGGGACG GTTTATTTTG CAGGACGGGT 1560

GTCAGGTGGT GGGGGAGCAG GGCGGCAGTC TGCACAATGT GACGCTGGAT GTTCGCGGGT 1620

CGGACTGTGT GATTAAAGGC GTGGCGATGA GCGGCTTTGG CCCCGTCGCG CAAATTTTCA 1680

TCGGTGGTAA GGAACCGCAG GTGATGCGTA ATCTCATTAT CGATGACATC ACCGTTACCC 1740

ACGCCAACTA CGCCATTCTC CGCCAGGGAT TTCATAACCA AATGGATGGC GCGCGGATTA 1800

CGCATAGCCG CTTTAGCGAT TTACAGGGGG ACGCCATTGA GTGGAATGTC GCGATTCACG 1860

ACCGCGACAT CCTGATTTCC GATCATGTCA TCGAACGCAT TAATTGTACC AATGGCAAAA 1920

TCAACTGGGG GATCGGCATC GGGCTGGCGG GTAGCACCTA TGACAACAGT TATCCTGAAG 1980

ACCAGGCAGT AAAAAACTTT GTGGTGGCCA ATATTACCGG ATCTGATTGC CGACAGCTTG 2040

TGCACGTAGA AAATGGCAAA CATTTCGTCA TTCGCAATGT CAAAGCCAAA AACATCACGC 2100

CCGGTTTCAG TAAAAATGCG GGTATTGATA ACGCAACGAT CGCAATTTAT GGCTGTGATA 2160

ATTTCGTCAT TGATAATATT GATATGACGA ATAGTGCCGG GATGCTCATC GGCTATGGCG 2220

TCGTTAAAGG AAAATACCTG TCAATTCCGC AAAACTTTAA ATTAAACGCT ATTCGGTTGG 2280

ATAATCGCCA GGTTGCTTAT AAATTACGCG GCATTCAAAT TTCCTCCGGC AACACCCCCT 2340

CTTTTGTCGC CATCACCAAT GTACGGATGA CGCGTGCTAC GCTGGAACTG CATAATCAAC 2400

CGCAGCACCT CTTTCTGCGC AATATCAACG TGATGCAAAC TTCAGCGATT GGCCCGGCGT 2460

TAAAAATGCA TTTCGATTTG CGTAAAGATG TACGTGGTCA ATTTATGGCC CGCCAGGACA 2520

CGCTGCTTTC CCTCGCTAAT GTTCATGCCA TCAATGAAAA CGGGCAGAGT TCCGTGGATA 2580

TCGACAGGAT TAATCACCAA ACCGTGAATG TCGAAGCAGT GAATTTTTCG CTGCCGAAGC 2640

GGGGAGGGTA AGTACCGCTA TTTTTACGAA AATTCCTGGG AAAAAGTTGT TCATACTTAA 2700

TGTTATGGTG CCGACTAAGA CGTAATGTAG AGCGTGCCAT CATTATCCCT GGCAGCAGAG 2760

TAATTCATGC TGGCGAAAAC AAGCTAAAGA GCTATAATTC AGCAACCATT TTACAGGTGG 2820

AAGAAACAAT GATGAATTTG AAAGCAGTTA TACCGGTAGC GGGTTTGGGT ATGCATATGT 2880

TGCCTGCCAC CAAGGCAATC CCAAAAGAGA TGCTACCGAT CGTCGACAAG CCAATGATTC 2940

AGTACATTGT CGATGAGATT GTGGCTGCAG GGATCAAAGA AATCGTGCTG GTGACTCACG 3000

CGTCTAAAAA CGCCGTTGAG AACCACTTCG ACACCTCTTA TGAACTTGAA TCACTTCTTG 3060

AGCAGCGCGT TAAGCGTCAG CTTTTGGCGG AAGTGCAATC TATCTGCCCA CCGGGCGTGA 3120

CGATTATGAA CGTTCGCCAG GCGCAGCCGT TAGGGCTGGG GCATTCTATT CTGTGCGCGC 3180

GTCCGGTCGT GGGCGATAAC CCTTTCATTG TGGTACTCCC GGATATTATT ATCGATGATG 3240

CTACCGCCGA TCCGCTGCGC TATAACCTTG CGGCGATGGT GGCGCGTTTC AATGAAACGG 3300

GTCGCAGCCA GGTGCTGGCG AAGCGCATGA AAGGTGATTT ATCGGAGTAT TCCGTTATCC 3360

AGACGAAAGA ACCTCTGGAT AATGAAGGCA AAGTCAGCCG GATTGTGGAG TTTATCGAAA 3420

AACCGGATCA GCCGCAGACG CTGGATTCCG ATTTGATGGC GGTAGGCCGT TATGTGCTTT 3480

CAGCCGACAT CTGGGCGGAA CTGGAAAGAA CCGAACCGGG CGCCTGGGGC CGCATCCAGC 3540

TCACCGATGC CATTGCTGAA CTGGCGAAAA AACAGTCGGT TGACGCGATG CTAATGACGG 3600

GTGACAGCTA TGACTGCGGT AAAAAAATGG GCTACATGCA GGCATTTGTG AAGTACGGGC 3660

TGCGCAACCT GAAAGAAGGA GCCAAGTTCC GTAAGAGCAT AGAGCAGCTT TTGCATGAAT 3720

AAGTATTAAC AACCGTGATA AATGGTTGGT GATAAACATA ATAACGGCAG TGAACATTCG 3780

AAGCGGCAAG TTGGCTGAAA CGAGTGTTGA CTGCCGTTTT AGTTTTGTAT AAAGGGCTTA 3840

AGTAACAAGG GGTTATCTGG AGCATTTTAA TGCTGATTTT ATAAGATTAA TCCTTGTTTC 3900

CGGATGCAAT TAATAAGACA ATTAGCGTTT AAGTTTTAGT GAGCTTTGCC CTGCTGGGCG 3960

AGGTTTGCAA CAAGTCGATA TGTACGCAGT GCACTGGTAG CTGATGAGCC AGGGGCGGTA 4020
GCGTGTGTAA CGACTTGAGC AATTAATTTT TATTGGCAAA TTAAATACCA CATTAAATAC 4080
GCCTTATGGA ATAGAAAAGT GAAGATACTT ATTACTGGCG GGGCAGGTTT TATTGGATCA 4140
GCTGTTGTCC GCCATATTAT TAAGAATACA CAGGACACTG TAGTTAATAT TGATAAATTA 4200
ACCTACGCCG GTAATCTTGA ATCCCTTTCT GATATTTCTG AAAGTAATCG CTACAATTTT 4260
GAACACGCGG ATATTTGTGA TTCCGCTGAA ATAACGCGTA TTTTTGAGCA GTACCAGCCG 4320
GACGCGGTGA TGCATTTGGC TGCGGAAAGT CATGTGGACC GTTCGATTAC CGGGCCAGCA 4380
GCATTTATTG AAACCAATAT CGTCGGCACC TATGCACTTC TTGAAGTTGC GCGTAAATAC 4440
TGGTCTGCCC TTGGCGAAGA TAAAAAAAAT AATTTTCGTT TTCATCATAT TTCCACTGAT 4500
GAAGTTTACG GCGATTTACC GCATCCTGAT GAAGTTGAAA ACAGCGTTAC GCTGCCGTTA 4560
TTTACTGAAA CGACGGCATA TGCGCCAAGT AGCCCCTATT CTGCGTCAAA AGCATCCAGC 4620
GATCATTTAG TCCGTGCCTG GCGGCGTACC TATGGTCTAC CAACGATCGT TACCAATTGT 4680
TCTAATAACT ATGGCCCTTA TCACTTCCCT GAAAAACTGA TTCCGTTGGT CATTTTGAAC 4740
GCACTGGAAG GAAAGCCTTT GCCAATTTAT GGCAAAGGGG ATCAGATTCG CGATTGGCTA 4800
TATGTAGAAG ATCATGCTCG CGCGCTTCAT ATGGTAGTGA CTGAAGGCAA GGCAGGGGAG 4860
ACTTATAACA TTGGTGGACA CAATGAGAAG AAAAATCTCG ATGTGGTATT TACCATCTGT 4920
GATCTGCTGG ATGAGATTGT ACCCAAAGCG ACTTCTTATC GTGAACAAAT CACTTATGTC 4980
GCGGATCGTC CGGGCCATGA TCGTCGTTAT GCCATTGATG CAGGTAAAAT TAGCCGCGAA 5040
TTAGGCTGGA AACCGCTGGA GACCTTTGAA AGCGGTATTC GTAAAACAGT GGAATGGTAC 5100
CTTGCAAATA CTCAATGGGT AAACAATGTT AAAAGTGGGG CGTATCAGAG TTGGATAGAA 5160
CAGAACTATG AAGGACGCCA GTAATGAATA TCTTACTTTT TGGTAAGACA GGGCAAGTAG 5220
GCTGGGAGTT GCAACGTTCT CTGGCACCGG TAGGGAATCT GATTGCCCTG GATGTCCATT 5280
CAAAAGAGTT TTGCGGTGAT TTTAGTAATC CGAAAGGCGT TGCCGAAACC GTTCGTAAGC 5340
TTCGTCCCGA TGTGATTGTT AACGCAGCAG CCCATACTGC AGTAGATAAA GCAGAGTCTG 5400
AACCAGAACT GGCGCAGTTA CTTAACGCCA CCAGTGTGGA AGCCATCGCT AAAGCAGCCA 5460
ACGAAACTGG CGCATGGGTA GTGCATTATT CAACCGATTA TGTATTTCCT GGTACCGGCG 5520
ATATCCCATG GCAGGAAACG GACGCTACGT CGCCGCTGAA TGTCTATGGC AAAACCAAAC 5580
TGGCGGGAGA AAAGGCCCTG CAGGATAACT GCCCTAAACA CCTTATCTTC CGCACCAGTT 5640
GGGTTTATGC AGGTAAGGGC AATAATTTCG CAAAGACAAT GCTTCGTCTG GCGAAAGAGC 5700
GTCAGACACT TTCAGTCATT AACGATCAGT ACGGTGCGCC AACCGGTGCG GAATTACTGG 5760
CTGACTGTAC GGCGCATGCG ATCCGTGTGG CGTTAAATAA ACCAGAAGTC GCAGGTCTTT 5820
ACCATCTGGT TGCCGGGGGA ACCACAACCT GGCATGACTA CGCGGCCTTA GTCTTTGACG 5880
AGGCGCGCAA AGCAGGGATA ACGCTTGCGC TGACTGAGCT TAATGCTGTG CCGACCAGCG 5940
CCTACCCGAC GCCGGCGAGC AGACCAGGCA ATTCGCGTCT CAATACTGAA AAGTTTCAGC 6000

GTAATTTTGA CCTTATTCTG CCTCAATGGG AATTAGGAGT TAAGCGTATG CTGACTGAAA 6060
TGTTTACGAC GACAACCATC TAATAAATTT AAATGCCCAT CAGGGCATTT TCTATGAATG 6120
AGAAATGGAA ATGAAAACGC GTAAGGGCAT TATTTTAGCG GGGGGCTCCG GCACCCGTCT 6180
TTATCCGGTG ACCATGGCGG TAAGTAAGCA ATTGCTACCA ATTTATGATA AACCGATGAT 6240
TTACTATCCC CTTTCCACGC TTATGCTGGC AGGCATTCGG GATATCCTGA TCATCAGTAC 6300
GCCACAGGAC ACGCCGCGTT TTCAACAACT GCTGGGAGAC GGCAGCCAGT GGGGGCTGAA 6360
TCTTCAATAT AAAGTACAGC CAAGCCCGGA TGGCTTAGCA CAGGCGTTTA TTATTGGTGA 6420
AGAGTTCATT GGTCATGATG ATTGTGCATT AGTGCTGGGT GACAATATCT TCTATGGTCA 6480
TGATTTACCA AAGTTAATGG AAGCTGCCGT TAATAAAGAA AGTGGTGCTA CCGTCTTCGC 6540
TTATCATGTA AACGATCCGG AGCGCTACGG TGTGGTTGAG TTTGACCAAA AGGGCACAGC 6600
CGTTAGTCTG GAAGAAAAAC CATTACAACC GAAGAGTAAT TACGCGGTAA CGGGGCTGTA 6660
TTTTTATGAT AATAGCGTGG TGGAGATGGC GAAAAATCTT AAGCCTTCCG CTCGCGGTGA 6720
GTTAGAAATC ACGGATATTA ACCGTATCTA TATGGAGCAG GGAAGATTGT CTGTCGCTAT 6780
GATGGGGCGC GGTTATGCCT GGCTGGATAC AGGGACGCAT CAGAGTTTGA TAGAGGCCAG 6840
TAATTTTATT GCAACCATCG AAGAACGCCA GGGGCTAAAA GTGTCCTGCC CGGAAGAGAT 6900
CGCATTTCGT AAAAATTTTA TAAATGCACA ACAGGTTATA GAACTGGCCG GCCATTATC 6960
AAAAAATGAT TATGGCAAAT ATTTGCTGAA GATGGTGAAA GGTTTATAAG TGATGATTGT 7020
GATTAAAACA GCAATACCAG ATGTCTTGAT CTTAGAGCCT AAAGTTTTTG GCGATGAGAG 7080
GGGATTCTTT TTTGAAAGTT ATAACCAGCA GACCTTTGAA GAGTTGATTG GACGTAAAGT 7140
TACATTTGTT CAAGATAATC ATTCAAAATC CAAAAAGAAC GTACTCAGAG GGCTACATTT 7200
TCAGAGAGGA GAAAATGCAC AGGGGAAGTT AGTTCGTTGT GCTGTCGGTG AGGTTTTTGA 7260
TGTTGCGGTC GATATCCGAA AAGAATCGCC TACTTTTGGT CAATGGGTTG GTGTAAATCT 7320
GTCTGCTGAG AATAAGCGAC AGCTTTGGAT TCCAGAAGGT TTTGCTCATG GTTTTGTTAC 7380
TCTTAGTGAG TATGCAGAGT TTCTGTACAA AGCAACTAAT TATTACTCAC CTTCATCGGA 7440
AGGTAGCATT CTATGGAATG ATGAGGCAAT AGGTATTGAA TGGCCTTTTT CTCAGCTGCC 7500
TGAGCTTTCA GCAAAAGATG CTGCAGCACC TTTACTGGAT CAAGCCTTGT TAACAGAGTA 7560
AGCATCGTGT CTCATATTAT TAAGATTTTT CCATCAAATA TTGAATTTTC CGGTAGAGAG 7620
GATGAATCAA TCCTCGATGC TGCGCTATCG GCTGGTATCC ATCTTGAACA TAGCTGCAAA 7680
GCGGGTGATT GTGGTATCTG TGAGTCCGAT TTGTTGGCGG GAGAAGTTGT TGACTCCAAA 7740
GGTAATATTT TTGGACAGGG TGATAAAATA CTAACCTGCT GCTGTAAACC TAAAACCGCC 7800
CTTGAGCTAA ATGCGCATTT TTTTCCTGAA CTAGCTGGAC AGACAAAAAA AATTGTCCCA 7860
TGCAAGGTAA ATAGTGCTGT ACTGGTTTCA GGCGATGTTA TGACTTTGAA GTTACGCACA 7920
CCACCAACAG CAAAAATTGG CTTCCTTCCA GGGCAGTATA TCAATTTACA TTATAAAGGT 7980
GTAACTCGCA GTTATTCTAT CGCTAATAGT GATGAGTCGA ATGGTATTGA GTTGCATGTA 8040

AGGAATGTTC CCAATGGTCA GATGAGTTCG CTCATTTTTG GGGAGTTACA AGAAAATACT 8100
CTTATGCGCA TTGAAGGGCC TTGCGGAACA TTTTTTATTC GTGAAAGTGA CAGACCTATA 8160
ATCTTCCTTG CAGGCGGTAC TGGATTCGCT CCAGTTAAAT CAATGGTTGA GCATCTCATT 8220
CAGGGAAAAT GTCGTCGTGA GATCTACATT TACTGGGGAA TGCAATATAG TAAAGATTTT 8280
TACTCTGCAT TACCGCAGCA GTGGAGTGAA CAGCACGACA ACGTTCATTA TATCCCTGTT 8340
GTTTCTGGTG ATGACGCCGA ATGGGGGGGA AGAAAGGGAT TTGTCCATCA TGCCGTGATG 8400
GATGATTTTG ATTCTCTAGA GTTCTTCGAT ATATATGCAT GTGGTTCACC TGTGATGATC 8460
GATGCCAGTA AAAAGGACTT TATGATGAAA AATCTCTCTG TAGAACATTT CTATTCTGAT 8520
GCATTTACCG CATCTAATAA TATTGAGGAT AATTTATGAA AGCGGTCATC CTGGCTGGTG 8580
GACTTGGTAC CAGACTAAGT GAAGAAACAA TTGTAAAACC AAAACCGATG GTAGAAATTG 8640
GTGGCAAGCC TATTCTTTGG CACATTATGA AAATGTATTC TGTGCATGGT ATCAAGGATT 8700
TTATTATCTG CTGTGGTTAT AAAGGATATG TGATTAAAGA ATATTTTGCG AACTACTTCC 8760
TTCACATGTC AGATGTAACA TTCCATATGG CTGAAAACCG TATGGAAGTT CACCATAAAC 8820
GTGTTGAACC ATGGAATGTC ACATTGGTTG ATACGGGTGA TTCTTCAATG ACTGGTGGTC 8880
GTCTGAAACG TGTTGCTGAA TACGTAAAAG ATGACGAGGC TTTCCTGTTT ACTTATGGTG 8940
ATGGCGTTGC CGACCTTGAT ATCAAAGCGA CTATCGATTT CCATAAGGCT CACGGTAAGA 9000
AAGCGACTTT AACAGCTACT TTTCCACCAG GACGCTTTGG CGCATTAGAT ATCCGAGCTG 9060
GTCAGGTCCG GTCATTCCAG GAAAAACCGA AAGGCGATGG GGCAATGATC AATGGTGGTT 9120
TCTTTGTGTT GAATCCATCG GTTATCGATC TCATCGATAA CGATGCAACA ACCTGGGAAC 9180
AAGAGCCATT AATGACATTG GCACAACAGG GGGAGTTAAT GGCTTTTGAA CACCCAGGTT 9240
TCTGGCAGCC GATGGATACC CTACGTGATA AAGTTTACCT CGAAGGGCTG TGGGAAAAAG 9300
GTAAAGCTCC GTGGAAAACC TGGGAGTAAC TAGATGATTG ATAAAAATTT TTGGCAAGGT 9360
AAACGTGTAT TCGTTACCGG CCATACTGGC TTTAAAGGAA GCTGGCTTTC GCTATGGCTG 9420
ACTGAAATGG GTGCAATTGT AAAAGGCTAT GCACTTGATG CGCCAACTGT TCCAAGTTTA 9480
TTTGAGATAG TGCGTCTTAA TGATCTTATG GAATCTCATA TTGGCGACAT TCGTGATTTT 9540
GAAAAGCTGC GCAATTCTAT TGCAGAATTT AAGCCAGAAA TTGTTTTCCA TATGGCAGCC 9600
CAGCCTTTAG TGCGCCTATC TTATGAACAG CCAATCGAAA CATACTCAAC AAATGTTATG 9660
GGTACTGTCC ATTTGCTTGA AACAGTTAAG CAAGTAGGTA ACATAAAGGC AGTCGTAAAT 9720
ATCACCAGTG ATAAGTGCTA CGACAATCGT GAGTGGGTGT GGGGCTATCG TGAGAACGAA 9780
CCCATGGGAG GGTACGATCC ATACTCTAAT AGTAAAGGTT GTGCAGAATT AGTCGCGTCT 9840
GCATTCCGGA ACTCATTCTT CAATCCTGCA AATTATGAGC AACATGGCGT TGGTTTGGCG 9900
TCTGTGAGGG CTGGTAATGT CATAGGCGGA GGCGATTGGG CTAAAGACCG TTTAATTCCC 9960
GATATTCTGC GCTCATTTGA AAATAACCAG CAGGTTATTA TTCGAAACCC ATATTCTATC 10020
CGTCCCTGGC AGCATGTACT GGAGCCTCTT TCTGGTTACA TTGTGGTGGC GCAACGCTTA 10080

TATACAGAAG GTGCTAAGTT TTCTGAAGGA TGGAATTTCG GCCCGCGTGA TGAAGATGCG 10140
AAGACGGTCG AATTTATTGT TGACAAGATG GTCACGCTTT GGGGTGATGA TGCAAGCTGG 10200
TTACTGGATG GTGAGAATCA TCCTCATGAG GCACATTACC TGAAACTGGA TTGCTCTAAA 10260
GCAAATATGC AATTAGGATG GCATCCGCGT TGGGGATTGA CTGAAACACT TGGTCGCATC 10320
GTAAAATGGC ATAAAGCATG GATTCGCGGC GAAGATATGT TGATTTGTTC AAAGCGTGAA 10380
ATCAGCGACT ATATGTCTGC AACTACTCGT TAAGAAAATA AGTTTAAGGA ATCAAAGTAA 10440
TGACAGCAAA TAACCTGCGT GAGCAAATCT CTCAGCTTGT CGCTCAGTAT GCGAATGAGG 10500
CATTGAGCCC GAAACCTTTT GTTGCAGGTA CAAGCGTTGT GCCTCCTTCC GGGAAGGTTA 10560
TTGGTGCCAA AGAGTTACAA TTGATGGTTG AGGCGTCTCT TGATGGATGG CTAACTACTG 10620
GTCGTTTCAA TGATGCCTTT GAAAAAAAAC TTGGGGAATT TATTGGGGTT CCTCATGTTT 10680
TAACGACAAC ATCTGGCTCT TCGGCAAACT TGCTGGCACT GACTGCGCTG ACTTCCCCAA 10740
AATTAGGCGA GCGAGCTCTC AAACCTGGTG ATGAGGTTAT TACTGTCGCT GCTGGCTTCC 10800
CGACTACAGT TAACCCGGCG ATCCAGAATG GTTTAATACC GGTATTCGTG GATGTTGATA 10860
TCCCGACATA TAATATCGAT GCCTCTCTCA TTGAAGCTGC AGTTACTGAG AAATCAAAAG 10920
CGATAATGAT CGCTCATACA CTCGGTAATG CATTTAACCT GAGTGAAGTT CGTCGGATTG 10980
CCGATAAATA TAACTTATGG TTGATTGAAG ACTGCTGTGA TGCCCTTGGG ACGACTTATG 11040
AAGGCCAGAT GGTAGGTACC TTTGGTGACA TCGGAACCGT TAGTTTTTAT CCGGCTCACC 11100
ATATCACAAT GGGTGAAGGC GGTGCTGTAT TCACCAAGTC AGGTGAACTG AAGAAAATTA 11160
TTGAGTCGTT CCGTGACTGG GGCCGGGATT GTTATTGTGC GCCAGGATGC GATAACACCT 11220
GCGGTAAACG TTTTGGTCAG CAATTGGGAT CACTTCCTCA AGGCTATGAT CACAAATATA 11280
CTTATTCCCA CCTCGGATAT AATCTCAAAA TCACGGACAT GCAGGCAGCA TGTGGTCTGG 11340
CTCAGTTGGA GCGCGTAGAA GAGTTTGTAG AGCAGCGTAA AGCTAACTTT TCCTATCTGA 11400
AACAGGGCTT GCAATCTTGC ACTGAATTCC TCGAATTACC AGAAGCAACA GAGAAATCAG 11460
ATCCATCCTG GTTTGGCTTC CCTATCACCC TGAAAGAAAC TAGCGGTGTT AACCGTGTCG 11520
AACTGGTGAA ATTCCTTGAT GAAGCAAAAA TCGGTACACG TTTACTGTTT GCTGGAAATC 11580
TGATTCGCCA ACCGTATTTT GCTAATGTGA AATATCGTGT AGTGGGTGAG TTGACAAATA 11640
CCGACCGTAT AATGAATCAA ACGTTCTGGA TTGGTATTTA TCCAGGCTTG ACTACAGAGC 11700
ATTTAGATTA TGTAGTTAGC AAGTTTGAAG AGTTCTTTGG TTTGAATTTC TAATTCAATT 11760
TATTCTATCT GGTGATTGCG ATGACCTTTT TGAAAGAATA TGTAATTGTC AGTGGGGCTT 11820
CCGGCTTTAT TGGTAAGCAT TTACTCGAAG CGCTAAAAAA ATCGGGGATT TCAGTTGTCG 11880
CAATCACTCG AGATGTAATA AAAAATAATA GTAATGCATT AGCTAATGTT AGATGGTGCA 11940
GTTGGGATAA TATCGAATTA TTAGTCGAGG AGTTATCAAT TGATTCTGCA TTAATTGGTA 12000
TCATTCATTT GGCAACAGAA TATGGGCATA AAACATCATC TCTCATAAAT ATTGAAGATG 12060
CAAATGTTAT AAAACCATTA AAGCTTCTTG ATTTGGCAAT AAAATATCGG GCGGATATCT 12120

TTTTAAATAC AGATAGTTTT TTTGCCAAGA AAGATTTTAA TTATCAACAT ATGCGGCCTT 12180

ATATAATTAC TAAAAGACAC TTTGATGAAA TTGGGCATTA TTATGCTAAT ATGCATGACA 12240

TTTCATTTGT AAACATGCGA TTAGAGCATG TATATGGGCC TGGGGATGGT GAAAATAAAT 12300

TTATTCCATA CATTATCGAC TGCTTAAATA AAAAACAGAG TTGCGTGAAA TGTACAACAG 12360

GCGAACAGAT AAGAGACTTT ATTTTTGTAG ATGATGTGGT AAATGCTTAT TTAACTATAT 12420

TAGAAAATAG AAAAGAAGTA CCTTCATATA CTGAGTATCA AGTTGGAACT GGTGCTGGGG 12480

TAAGTTTGAA AGATTTTCTG GTTTATTTGC AAAATACTAT GATGCCAGGT TCATCGAGTA 12540

TATTTGAATT TGGTGCGATA GAGCAAAGAG ATAATGAAAT AATGTTCTCT GTAGCAAATA 12600

ATAAAAATTT AAAAGCAATG GGCTGGAAAC CAAATTTCGA TTATAAAAAA GGAATTGAAG 12660

AACTACTGAA ACGGTTATGA GATTTTCATG ATCTTTTAAT AAATAAATCG TTAACAAATT 12720

AGTCGCGTTA TGTTGTAAAA ACTAAGTCGT TTAATTGCAT AGTGAAAGTT CAATTGTTAA 12780

AAATTCCGAG TCATTTAATT GTTGCAGGTT CATCATGGTT ATCCAAAATA ATAATTGCCG 12840

GGGTGCAGTT AGCAAGTATT TCATATCTTA TTTCTATGCT AGGTGAAGAG AAATATGCAA 12900

TCTTTAGTTT GTTAACTGGT TTATTAGTAT GGTGTAGCGC TGTTGATTTT GGCATAGGTA 12960

CAGGACTGCA AAATTATATA TCAGAATGCA GAGCCAAAAA CAAAAGTTAT GATGCATATA 13020

TTAAATCAGC ATTACATCTA AGCTTTATAG CTATTATTTT TTTTATTGCT TTATTTTATA 13080

TTTTTTCTGG GGTAATTTCC GCTAAATATC TTTCTTCTTT TCATGAGGTA TTACAGGACA 13140

AAACCAGAAT GCTCTTTTTT ACCTCATGTC TGGTTTTCAG TTCTATTGGA ATCGGAGCTA 13200

TTGCTTATAA AATACTTTTT GCCGAATTGG TCGGGTGGAA AGCTAATCTA TTAAACGCAT 13260

TATCTTATAT GATAGGTATG CTCGGCTTGC TATATATATA CTATAGGGGG ATCTCAGTTG 13320

ACATAAAATT ATCACTAATA GTCCTGTATC TTCCAGTGGG TATGATTTCA TTGTGCTATA 13380

TTGTATATAG ATACATAAAG CTTTATCATG TTAAAACAAC AAAATCTCAT TATATAGCAA 13440

TTTTACGTAG ATCTTCAGGG TTTTTTCTTT TTACTTTATT ATCGATAGTG GTGCTTCAAA 13500

CAGATTATAT GGTCATTTCT CAAAGGCTAA CTCCTGCTGA TATTGTTCAA TATACAGTAA 13560

CGATGAAAAT TTTTGGTTTA GTCTTTTTTA TTTATACTGC TATTTTGCAA GCATTATGGC 13620

CTATATGTGC TGAATTGAGA GTCAAACAGC AATGGAAAAA ACTTAACAAA ATGATAGGTG 13680

TCAATATTTT GCTTGGCTCA CTATATGTTG TTGGATGTAC AATATTTATT TATTTATTTA 13740

AAGAACAGAT ATTTTCAGTA ATAGCCAAAG ATATTAATTA TCAAGTTTCT ATTTTATCTT 13800

TTATGTTAAT TGGCATATAT TTCTGTATTC GCGTTTGGTG TGACACTTAT GCAATGTTAT 13860

TGCAAAGTAT GAATTATTTA AAAATACTTT GGATATTAGT ACCACTACAA GCAATAATTG 13920

GTGGAATAGC ACAATGGTAT TTTTCTAGTA CGCTTGGAAT CAGTGGAGTG CTGCTTGGCT 13980

TGATTATATC TTTTGCTTTA ACTGTTTTTT GGGGGCTTCC ACTAACTTAC TTAATTAAGG 14040

CAAATAAGGG ATAATCATAT GCTTATATCA TTTTGTATTC CAACTTATAA TAGAAAACAA 14100

TATCTTGAAG AGTTGTTGAA TAGTATAAAT AATCAGGAAA AATTTAATTT AGATATTGAG 14160

ATATGTATAT CAGATAATGC CTCTACTGAT GGTACAGAGG AAATGATTGA TGTTTGGAGG 14220
AACAATTATA ATTTCCCAAT AATATATCGG CGTAATAGCG TTAACCTTGG GCCAGATAGG 14280
AATTTTCTTG CTTCAGTATC CCTTGCGAAT GGGGATTATT GTTGGATATT TGGCAGTGAT 14340
GATGCTCTTG CGAAAGACTC GTTAGCGATA TTACAAACTT ATCTCGATTC TCAAGCAGAT 14400
ATATATTTAT GTGACAGAAA AGAGACCGGG TGTGATTTAG TTGAGATTAG AAACCCTCAT 14460
CGTTCTTGGC TCAGAACAGA TGATGAACTT TATGTGTTTA ATAATAATTT AGATAGGGAA 14520
ATCTATCTCA GTAGATGCTT ATCTATTGGT GGTGTATTTA GCTATCTAAG TTCTTTAATA 14580
GTAAAAAAAG AACGATGGGA TGCCATTGAT TTTGATGCGT CCTATATTGG CACTTCCTAT 14640
CCTCATGTAT TTATCATGAT GAGCGTATTT AATACGCCAG GGTGCCTTTT GCATTATATA 14700
TCAAAACCAC TCGTAATATG CCGAGGAGAT AATGATAGTT TCGAGAAGAA AGGAAAGGCC 14760
AGACGAATTT TAATTGATTT TATTGCATAT TTAAAATTAG CTAATGATTT TTACAGTAAA 14820
AATATATCTT TAAAACGAGC ATTTGAAAAT GTTTTGCTAA AAGAGAGACC ATGGTTATAT 14880
ACAACTTTGG CTATGGCATG TTATGGCAAT AGTGATGAAA AAAGAGATTT ATCTGAATTT 14940
TATGCAAAGC TAGGTTGTAA TAAAAATATG ATCAACACTG TACTTCGATT TGGGAAACTA 15000
GCATATGCAG TGAAAAATAT TACCGTGCTT AAGAATTTTA CTAAACGGAT AATTAAGTAG 15060
TAGTAAGTTA TTATATTGAG ATTAAATGTA GATTTAACCT TTCTGGATTC AGCTAGATTT 15120
ACGTTACTGA CTTTTCTTTT TAATGAAAAT CATATTTGAT ATATATAAAT AAATTTGGAT 15180
AGCTTAACTA CTTAGATGTT TTTTCTGGG AATGTTAGTA TAATAATATA TTTCTTTATG 15240
ATTGTTTTTG TAGTGTTTTA CTGCCGGTAT TACATTAACT CTATTATTAA GAATTACACC 15300
TAGTGTAAGC TTCGTAATAT TATTTATCCT TATGATTATT GCTTTAAAGA TGCGTATGGA 15360
AAAACGGAGA GCTATTCAAT GATCGTAAAC CTATCACGTT TAGGTAAAAG TGGTACGGGA 15420
ATGTGGCAAT ACTCGATTAA ATTTTTAACG GCACTGCGAG AAATAGCTGA TGTTGACGCA 15480
ATAATCTGTA GCAAGGTACA CGCTGATTAT TTTGAAAAGC TCGGTTATGC AGTAGTTACT 15540
GTTCCGAATA TTGTTAGCAA CACATCAAAA ACATCGCGAC TTAGACCATT AGTATGGTAT 15600
GTATATAGTT ACTGGCTTGC GCTGAGGGTT TTAATTAAGT TTGGTAATAA AAAATTGGTG 15660
TGTACTACAC ATCACACTAT CCCCTTACTG AGAAACCAAA CGATAACCGT ACATGATATA 15720
AGACCTTTTT ATTATCCAGA TAGTTTTATT CAGAAAGTGT ATTTTCGCTT TTTATTAAAA 15780
ATGTCCGTTA AGCGATGTAA GCATGTTTTA ACGGTATCTT ATACCGTTAA AGATAGCATT 15840
GCTAAAACTT ATAATGTAGA TAGTGAGAAA ATATCAGTAA TTTATAATAG TGTTAATAAA 15900
TCTGATTTTA TACAAAAAAA AGAAAAAGAG AATTACTTTT TAGCTGTTGG TGCAAGTTGG 15960
CCACATAAAA ATATTCATTC ATTCATAAAA AATAAAAAAG TTTGGTCTGA CTCTTATAAT 16020
TTAATTATTG TATGTGGTCG TACTGACTAT GCAATGTCTC TCCAACAAAT GGTCGTTGAT 16080
CTGGAACTAA AAGATAAAGT GACTTTTTTA CATGAAGTCT CATTTAATGA ATTAAAGATT 16140
TTATATTCTA AAGCCTACGC GCTTGTTTAT CCATCTATTG ATGAGGGTTT TGGTATACCT 16200

CCTATTGAAG CGATGGCATC AAATACTCCA GTTATAGTGT CCGATATACC AGTATTTCAT 16260
GAAGTGTTAA CCAATGGTGC ATTATATGTG AATCCGGATG ATGAAAAAAG CTGGCAGAGT 16320
GCAATTAAAA ATATAGAGCA GTTGCCTGAT GCAATTTCCC GATTTAACAA CTATGTCGCA 16380
CGGTATGACT TTGATAATAT GAAGCAGATG GTTGGCAATT GGTTGGCGGA ATCAAAATAA 16440
ATGAAAATAA CATTAATTAT TCCCACATAT AATGCAGGGT CGCTTTGGCC TAATGTTCTG 16500
GATGCGATTA AGCAGCAAAC TATATATCCG GATAAATTGA TTGTTATAGA CTCAGGTTCT 16560
AAAGATGAAA CGGTTCCGTT AGCCTCAGAC CTGAAAAATA TATCAATATT TAATATTGAC 16620
TCTAAAGATT TTAATCATGG AGGAACCAGA AATTTAGCAG TTGCAAAAAC TCTGGACGCT 16680
GATGTTATAA TTTTTCTAAC GCAAGATGCA ATTCTCGCGG ATTCGGATGC AATTAAAAAT 16740
TTGGTTTATT ATTTTTCAGA TCCATTGATA GCAGCGGTTT GTGGTAGACA ACTTCCTCAT 16800
AAAGATGCTA ATCCCCTTGC AGTGCATGCC AGAAATTTTA ATTATAGTTC AAAATCTATT 16860
GTTAAAAGTA AGGCAGATAT AGAAAAATTG GGTATTAAAA CTGTATTTAT GTCCAATTCT 16920
TTTGCTGCCT ATCGCCGTTC CGTTTTTGAA GAGTTAAGTG GGTTTCCTGA ACATACAATT 16980
CTTGCCGAGG ATATGTTTAT GGCGGCTAAG ATGATTCAGG CGGGTTATAA GGTCGCCTAC 17040
TGCGCTGAAG CGGTGGTAAG ACACTCCCAT AATTATACCC CGCGAGAAGA GTTTCAACGA 17100
TATTTTGATA CTGGTGTATT TCATGCTTGT TCTCCGTGGA TTCAGCGTGA CTTTGGCGGA 17160
GCCGGTGGTG AGGGTTTCCG CTTCGTAAAA TCAGAGATTC AATTCCTGCT TAAAAATGCA 17220
CCGTTCTGGA TTCCAAGAGC TTTATTAACA ACCTTTGCTA AATTCTTGGG TTACAAATTA 17280
GGCAAGCATT GGCAATCTTT ACCGTTGTCT ACATGTCGCT ATTTTAGCAT GTACAAGAGT 17340
TATTGGAATA ATATCCAATA TTCTTCGTCA AAAGAGATAA AATAAATGTC TTTTCTTCCC 17400
GTAATTATGG CTGGCGGCAC AGGTAGCCGT TTATGGCCGC TTTCACGCGA ATATCATCCG 17460
AAGCAGTTTC TAAGCGTTGA AGGTAAACTA TCAATGCTGC AAAATACTAT AAAGCGATTA 17520
GCTTCACTTT CTACAGAAGA ACCCGTTGTC ATTTGCAATG ACAGACACCG TTTCTTAGTC 17580
GCTGAACAAC TCCGTGAAAT TGACAAGTTA GCAAATAATA TTATTCTCGA ACCGGTAGGC 17640
CGTAATACTG CACCAGCGAT CGCTCTTGCC GCGTTTTGTG CGCTCCAGAA TGCTGATAAT 17700
GCTGATCCTC TTTTGTTGGT TCTTGCTGCA GATCATGTGA TTCAGGATGA AATAGCTTTT 17760
ACGAAAGCTG TCAGACATGC TGAAGAATAC GCTGCAAATG GTAAGCTTGT AACTTTTGGT 17820
ATTGTTCCAA CGCATGCTGA AACGGGTTAT GGATATATTC GTCGTGGTGA GTTGATAGGA 17880
AATGACGCTT ATGCAGTGGC TGAATTTGTG GAGAAACCGG ATATCGATAC CGCCGGTGAC 17940
TATTTCAAAT CAGGGAAATA TTACTGGAAT AGCGGTATGT TTTTATTTCG TGCAAGCTCT 18000
TATTTAAACG AATTAAAGTA TTTATCACCT GAAATTTATA AAGCTTGTGA AAAGGCGGTA 18060
GGACATATAA ATCCCGATCT TGATTTTATT CGTATTGATA AAGAAGAGTT TATGTCATGC 18120
CCGAGTGATT CTATCGATTA TGCAGTTATG GAGCACACAC AGCATGCGGT GGTGATACCA 18180
ATGAGCGCTG GCTGGTCGGA TGTGGGTTCC TGGTCCTCAC TTTGGGATAT ATCGAATAAA 18240

GATCATCAGA GAAATGTTTT AAAAGGAGAT ATTTTCGCAC ATGCTTGTAA TGATAATTAC 18300

ATTTATTCCG AAGATATGTT TATAAGTGCG ATTGGTGTAA GCAATCTTGT CATTGTTCAA 18360

ACAACAGACG CTTTACTGGT GGCTAATAAA GATACAGTAC AAGATGTTAA AAAAATTGTC 18420

GATTATTTAA AACGGAATGA TAGGAACGAA TATAAACAAC ATCAAGAAGT TTTCCGCCCC 18480

TGGGGAAAAT ATAATGTGAT TGATAGCGGC AAAAATTACC TCGTTCGATG TATCACTGTT 18540

AAGCCGGGTG AGAAATTTGT GGCGCAGATG CATCACCACC GGGCTGAGCA TTGGATAGTA 18600

TTATCCGGGA CTGCTCGTGT TACAAAGGGA GAGCAGACTT ATATGGTTTC TGAAAATGAA 18660

TCAACATTTA TTCCTCCGAA TACTATTCAC GCGCTGGAAA ATCCTGGAAT GACCCCCCTG 18720

AAGTTAATTG AGATTCAATC AGGTACCTAT CTTGGTGAGG ATGATATTAT TCGTTTAGAA 18780

CAACGTTCTG GATTTTCGAA GGAGTGGACT AATGAACGTA GTTAATAATA GCCGTGATGT 18840

TATTTATTCA TCAGGTATTG TGTTTGGAAC GAGTGGGGCT CGCGGTCTTG TAAAAGATTT 18900

TACACCTCAG GTATGTGCTG CTTTTACGGT TTCATTTGTT GCCGTTATGC AGGAACATTT 18960

TTCCTTTGAT ACCGTAGCAT TGGCAATAGA TAATCGTCCA AGTAGTTATG GGATGGCTCA 19020

GGCGTGTGCT GCTGCATTGG CGGATAAAGG CGTTAACTGT ATTTTTTATG GAGTGGTACC 19080

AACCCCAGCT TTGGCCTTTC AGTCTATGTC TGACAATATG CCTGCGATAA TGGTTACGGG 19140

AAGTCATATT CCATTCGAGC GGAACGGCCT CAAGTTTTAT CGTCCTGATG GTGAAATCAC 19200

GAAACATGAT GAGGCTGCGA TCCTTAGTGT TGAAGATACG TGCAGCCATT TAGAGCTTAA 19260

AGAACTCATA GTTTCAGAAA TGGCTGCTGT TAATTATATA TCTCGTTATA CATCTTTATT 19320

TTCTACTCCA TTCCTGAAAA ATAAGCGTAT TGGTATTTAC GAACATTCAA GCGCTGGGCG 19380

TGATCTTTAT AAGCCTTTAT TTATTGCATT GGGGGCTGAA GTCGTTAGCT TGGGTAGAAG 19440

CGATAATTTT GTACCTATAG ATACAGAGGC TGTAAGCAAA GAGGATCGGG AAAAAGCTCG 19500

CTCATGGGCT AAAGAGTTCG ATTTACATGC CATATTCTCG ACAGATGGGG ATGGTGATCG 19560

CCCTCTTATT GCTGATGAGG CCGGTGAGTG GCTAAGAGGC GATATACTAG GTCTATTATG 19620

TTCACTTGCA TTGGATGCAG AAGCCGTCGC TATTCCTGTT AGTTGTAACA GCATAATTTC 19680

TTCTGGCCGC TTTTTTAAAC ATGTTAAGCT TACAAAAATT GGCTCGCCTT ATGTTATCGA 19740

AGCTTTTAAT GAATTATCGC GGAGTTATAG TCGTATTGTC GGTTTTGAAG CCAATGGCGG 19800

TTTTTTATTA GGAAGCGACA TCTGTATTAA CGAGCAGAAT CTTCATGCCT TACCAACTCG 19860

TGATGCTGTA TTACCAGCAA TAATGCTGCT TTACAAAAGT AGGAATACCA GCATTAGCGC 19920

TTTAGTCAAT GAACTCCCAA CTCGTTACAC CCATTCTGAC AGATTACAGG GGATTACAAC 19980

TGATAAAAGT CAATCCTTAA TTAGTATGGG CAGAGAAAAT CTGAGCAACC TCTTAAGCTA 20040

TATTGGTTTG GAGAATGAAG GTGCAATTTC TACAGATATG ACAGATGGTA TGCGAATTAC 20100

TTTACGTGAT GGATGTATTG TGCATTTGCG CGCTTCTGGT AATGCACCTG AGTTACGCTG 20160

CTATGCAGAA GCTAATTTAT TAAATAGGGC TCAGGATCTT GTAAATACAA CGCTTGCTAA 20220

TATTAAAAAA CGATGCTTGC TGTAAAAAAA TTGAATGTTA TTTACTTAAT ATGCCTATTT 20280

TATTTACATT ATGCACGGTC AGAGGGTGAG GATTAAATGG ATAATATTGA TAATAAGTAT 20340

AATCCACAGC TATGTAAAAT TTTTTTGGCT ATATCGGATT TGATTTTTTT TAATTTAGCC 20400

TTATGGTTTT CATTAGGATG TGTCTATTTT ATTTTTGATC AAGTACAGCG ATTTATTCCT 20460

CAAGACCAAT TAGATACAAG AGTTATTACG CATTTTATTT TGTCAGTAGT ATGTGTCGGT 20520

TGGTTTTGGA TTCGTTTGCG ACATTATACT ATCCGCAAGC CATTTTGGTA TGAGTTAAAA 20580

GAAATTTTTC GTACGATCGT TATTTTTGCT ATATTTGATT TGGCTCTGAT AGCGTTTACA 20640

AAATGGCAGT TTTCACGCTA TGTCTGGGTG TTTTGTTGGA CTTTTGCCCT AATCCTGGTG 20700

CCTTTTTTTC GCGCACTTAC AAAGCATTTA TTGAACAAGC TAGGTATCTG GAAGAAAAAA 20760

ACTATCATCC TGGGGAGCGG ACAGAATGCT CGTGGTGCAT ATTCTGCGCT GCAAAGTGAG 20820

GAGATGATGG GGTTTGATGT TATCGCTTTT TTTGATACGG ATGCGTCAGA TGCTGAAATA 20880

AATATGTTGC CGGTGATAAA GGATACTGAG ATTATTTGGG ATTTAAATCG TACAGGTGAT 20940

GTCCATTATA TCCTTGCTTA TGAATACACC GAGTTGGAGA AAACACATTT TTGGCTACGT 21000

GAACTTTCAA AACATCATTG TCGTTCTGTT ACTGTAGTCC CCTCGTTTAG AGGATTGCCA 21060

TTATATAATA CTGATATGTC TTTTATCTTT AGCCATGAAG TTATGTTATT AAGGATACAA 21120

AATAACTTGG CTAAAAGGTC GTCCCGTTTT CTCAAACGGA CATTTGATAT TGTTTGTTCA 21180

ATAATGATTC TTATAATTGC ATCACCACTT ATGATTTATC TGTGGTATAA AGTTACTCGA 21240

GATGGTGGTC CGGCTATTTA TGGTCACCAG CGAGTAGGTC GGCATGGAAA ACTTTTTCCA 21300

TGCTACAAAT TTCGTTCTAT GGTTATGAAT TCTCAAGAGG TACTAAAAGA ACTTTTGGCT 21360

AACGATCCTA TTGCCAGGGC TGAATGGGAG AAAGATTTTA AACTGAAAAA TGATCCTCGA 21420

ATCACAGCTG TAGGTCGATT TATACGTAAA ACTAGCCTTG ATGAGTTGCC ACAACTTTTT 21480

AATGTACTAA AAGGTGATAT GAGCCTGGTT GGACCACGAC CTATCGTTTC GGATGAACTG 21540

GAGCGTTATT GTGATGATGT TGATTATTAT TTGATGGCAA AGCCGGGCAT GACAGGTCTA 21600

TGGCAAGTGA GTGGGCGTAA TGATGTTGAT TATGACACTC GTGTTTATTT TGATTCCTGG 21660

TATGTTAAAA ACTGGACGCT TTGGAATGAT ATTGCCATTC TGTTTAAAAC AGCGAAAGTT 21720

GTTTTGCGGC GAGATGGTGC GTATTAAGCT TACCGAGAAG TACTGAATAA TAATTGTATA 21780

AATTAGCCTG CGTAAAATCT GAACGCATCA ATCGCTACCT TAATATCATA CCTTTGAGTT 21840

AACATACTAT TCACCTTTAA CCTGCCATGA CCGTTTGTGG CAGGGTTTCC ACACCTGACA 21900

GGAGTATGTA ATGTCCAAGC AACAGATCGG CGTCGTCGGT ATGGCAGTGA TGGGGCGCAA 21960

CCTCGCGCTC AACATCGAAA GCCGTGGTTA TACCGTCTCC GTTTTCAACC GCTCCCGTGA 22020

AAAGACCGAA GAAGTGATTG CCGAGAATCC CGGCAAAAAG CTGGTGCCTT ATTACACGGT 22080

THE CLAIMS:

1. A nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, including a *wzx* gene or a *wzy* gene, or a gene with a similar function; the gene being involved in the synthesis of a particular bacterial polysaccharide antigen, wherein the sequence of the nucleic acid molecule is specific to the particular bacterial polysaccharide antigen.

2. A nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit such as a *wzx* or *wzy* gene; the gene being involved in the synthesis of a particular bacterial O antigen, wherein the sequence of the nucleic acid molecule is specific to the particular bacterial O antigen.

3. A nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit such as a *wzx* or *wzy* gene; the gene being involved in the synthesis of an O antigen expressed by E. coli, wherein the sequence of the nucleic acid molecule is specific to the O antigen.

4. A nucleic acid molecule derived from a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit such as a *wzx* or *wzy* gene; the gene being involved in the synthesis of an O antigen expressed by S. enterica, wherein the sequence of the nucleic acid molecule is specific to the O antigen.

5. A nucleic acid molecule according to any one of claims 1 to 4 wherein the nucleic acid molecule is

approximately 10 to 20 nucleotides in length.

6. A nucleic acid molecule derived from a gene, the gene being selected from a group consisting of the following sequences:
nucleotide position 739 to 1932 of SEQ ID NO:1;
nucleotide position 8646 to 9911 of SEQ ID NO:1;
nucleotide position 9901 to 10953 of SEQ ID NO:1;
nucleotide position 11821 to 12945 of SEQ ID NO:1;
nucleotide position 79 to 861 of SEQ ID NO:2;
nucleotide position 858 to 2042 of SEQ ID NO:2;
nucleotide position 2011 to 2757 of SEQ ID NO:2;
nucleotide position 2744 to 4135 of SEQ ID NO:2;
nucleotide position 5257 to 6471 of SEQ ID NO:2; and
nucleotide position 13156 to 13821 of SEQ ID NO:2;
which nucleic acid molecule is capable of hybridizing to complementary sequence from said gene.

7. A nucleic acid molecule which is any one of the oligonucleotides in Table 5 or 5A, with respect to the genes *wbdH*, *wzx*, *wzy* and *wbdM*.

8. A nucleic acid molecule which is any one of the oligonucleotides in Table 6 or 6A.

9. A nucleic acid molecule derived from a gene, the gene being selected from a group consisting of the following sequences:
nucleotide position 1019 to 2359 of SEQ ID NO:3;
nucleotide position 2352 to 3314 of SEQ ID NO:3;
nucleotide position 3361 to 3875 of SEQ ID NO:3;
nucleotide position 3977 to 5020 of SEQ ID NO:3;
nucleotide position 5114 to 6313 of SEQ ID NO:3;
nucleotide position 6313 to 7323 of SEQ ID NO:3;
nucleotide position 7310 to 8467 of SEQ ID NO:3;
nucleotide position 12762 to 14054 of SEQ ID NO:4; and
nucleotide position 14059 to 15060 of SEQ ID NO:4;
which nucleic acid molecule is capable of hybridizing to

complementary sequences from said gene.

10. A nucleic acid molecule which is any one of the oligonucleotides in Table 7.

11. A nucleic acid molecule which is any one of the oligonucleotides in Table 8 with respect to the genes *wzx* and *wbaV*.

12. A method of testing a sample for the presence of one or more bacterial polysaccharide antigens, the method comprising the following steps:
(a) contacting the sample with at least one oligonucleotide molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing of oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the bacterial polysaccharide antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the bacterial polysaccharide antigen present in the sample and
(b) detecting any specifically hybridised oligonucleotide molecules.

13. The method according to claim 12, the method further comprising contacting the sample with a further at least one oligonucleotide molecule capable of specifically hybridising to at least one sugar pathway gene under conditions suitable to permit the further at least one oligonucleotide molecule to specifically hybridise to at least one such sugar pathway gene of any bacteria expressing the bacterial polysaccharide antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

14. A method of testing a sample for the presence

of one or more bacterial polysaccharide antigens, the method comprising the following steps:
(a) contacting the sample with at least one pair of oligonucleotide molecules, with at least one oligonucleotide molecule of the pair capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing of oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene; wherein the gene is involved in the synthesis of the bacterial polysaccharide antigen; under conditions suitable to permit the at least one oligonucleotide molecule of the pair of molecules to specifically hybridise to at least such gene of any bacteria expressing the bacterial polysaccharide antigen present in the sample and
(b) detecting any specifically hybridised oligonucleotide molecules.

15. The method according to claim 14, the method further comprising contacting the sample with a further at least one pair of oligonucleotide molecules, with at least one oligonucleotide molecule of the pair capable of specifically hybridising to at least one sugar pathway gene under conditions suitable to permit the further at least one oligonucleotide molecule of the pair to specifically hybridise to at least one such sugar pathway gene of any bacteria expressing the bacterial polysaccharide antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

16. A method of testing a sample for the presence of one or more bacterial O antigens, the method comprising the following steps:
(a) contacting the sample with at least one oligonucleotide molecule capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or

polysaccharide units, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the bacterial O antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the bacterial O antigen present in the sample and

(b) detecting any specifically hybridised oligonucleotide molecules.

17. The method according to claim 16, the method further comprising contacting the sample with a further at least one oligonucleotide molecule capable of specifically hybridising to at least one sugar pathway gene under conditions suitable to permit the further at least one oligonucleotide molecule to specifically hybridise to at least one such sugar pathway gene of any bacteria expressing the bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

18. The method according to claim 16 or 17 wherein the O antigen is expressed by E. coli or S. enterica.

19. The method according to claim 18 wherein the E. coli express the 0157 O antigen serotype or the 0111 O antigen serotype.

20. The method according to claim 18 wherein the S. enterica express the C2 or B O antigen serotype.

21. The method according to any one of claims 16 to 20 wherein the specifically hybridised oligonucleotide molecules are detected by Southern blot analysis.

22. A method of testing a sample for the presence of one or more bacterial O antigens, the method comprising the following steps:

(a) contacting the sample with at least one pair of oligonucleotide molecules, with at least one oligonucleotide molecule of the pair being capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene; wherein the gene is involved in the synthesis of the bacterial O antigen; under conditions suitable to permit the at least one oligonucleotide molecule of the pair of molecules to specifically hybridise to at least one such gene of any bacteria expressing the bacterial O antigen present in the sample and

(b) detecting any specifically hybridised oligonucleotide molecules.

23. The method according to claim 22, the method further comprising contacting the sample with a further at least one pair of oligonucleotide molecules, with at least one oligonucleotide molecule of the pair capable of specifically hybridising to at least one sugar pathway gene under conditions suitable to permit the further at least one oligonucleotide molecule of the pair to specifically hybridise to at least one such sugar pathway gene of any bacteria expressing the bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

24. The method according to claim 22 or 23 wherein the O antigen is expressed by E. coli or S. enterica.

25. The method according to claim 24 wherein the E. coli are 0111 or the 0157 O antigen serotype.

26. The method according to claim 24 wherein the S. enterica express the C2 or B O antigen serotype.

27. The method according to any one of claims 22 to 26 wherein the method is performed according to the polymerase chain reaction method.

28. The method according to any one of claims 22 to 26 wherein the oligonucleotide molecules are selected from the group of nucleic acid molecules according to any one of claims 5 to 11.

29. A method for testing a food derived sample for the presence of one or more particular bacterial O antigens, the method being according to any one of claims 16 to 28.

30. A method for testing a faecal derived sample for the presence of one or more particular bacterial O antigens, the method being according to any one of claims 16 to 28.

31. A method for testing a sample derived from a patient for the presence of one or more particular bacterial O antigens, the method being according to any one of claims 16 to 28.

32. A kit comprising a first vial containing a first nucleic acid molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene, wherein said gene is involved in the synthesis of a bacterial polysaccharide.

33. The kit according to claim 32 further comprising in the first vial, or in a second vial, a second nucleic acid molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene, wherein

said gene is involved in the synthesis of a bacterial polysaccharide, and wherein the sequence of the second nucleic acid molecule is different from the sequence of the first nucleic acid molecule.

34. The kit according to claim 33 further comprising a nucleic acid molecule derived from a sugar pathway gene.

35. A kit according to claim 32 further comprising in the first vial, or in a second vial, a second nucleic acid molecule capable of specifically hybridising to a sugar pathway gene.

36. A kit according to any one of claims 32 to 35 wherein the nucleic acid molecules are approximately 10 to 20 nucleotides in length.

37. A kit comprising a first vial containing a first nucleic acid molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene, wherein said gene is involved in the synthesis of a bacterial O antigen.

38. The kit according to claim 37, further comprising in the first vial, or in a second vial, a second nucleic acid molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene, wherein said gene is involved in the synthesis of a bacterial O antigen, and wherein the sequence of the second nucleic acid molecule is different from the sequence of the first nucleic acid molecule.

39. A kit according to claim 37 further comprising in the first vial, or in a second vial, a second nucleic acid molecule capable of specifically hybridising to a

sugar pathway gene.

40. The kit according to claim 38 further comprising a nucleic acid molecule derived from a sugar pathway gene.

41. The kit according to any one of claims 37 to 40 wherein the nucleic acid molecules are approximately 10 to 20 nucleotides in length.

42. The kit according to any one of claims 31 to 34 wherein the first and second nucleic acid molecules are according to any one of claims 5 to 11.

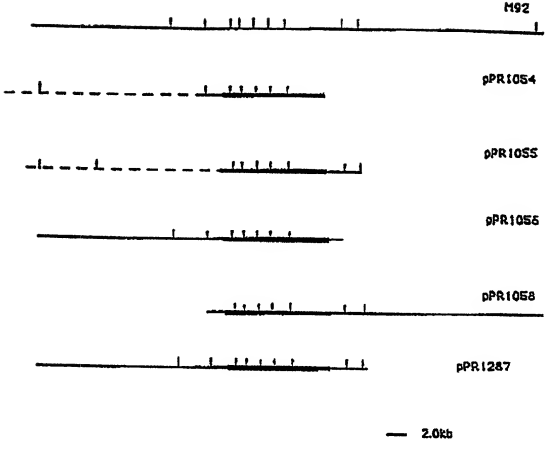

Figure 1

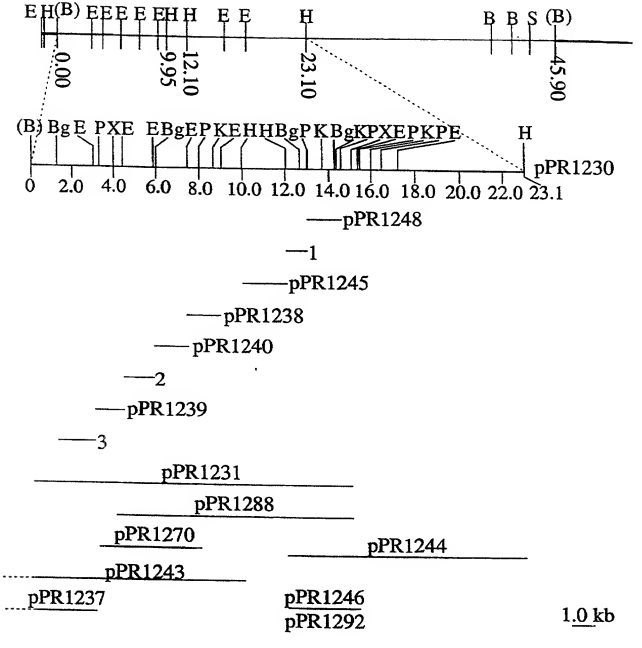
E H(B) E E E E E H H E E H B B S (B)
0.00
9.95
12.10
23.10
45.90
(B) Bg E PXE EBgEPKEHHBgPKBgKPXEPKPE H
pPR1230
0 2.0 4.0 6.0 8.0 10.0 12.0 14.0 16.0 18.0 20.0 22.0 23.1
pPR1248
1
pPR1245
pPR1238
pPR1240
2
pPR1239
3
pPR1231
pPR1288
pPR1270
pPR1244
pPR1243
pPR1237
pPR1246
pPR1292
1.0 kb

Figure 2

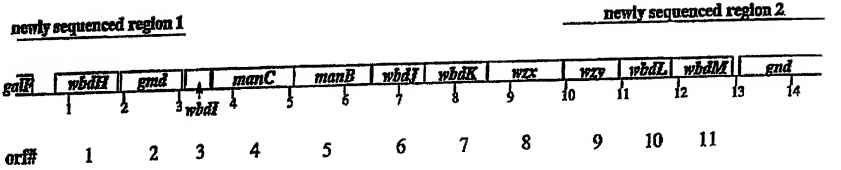

Figure 3

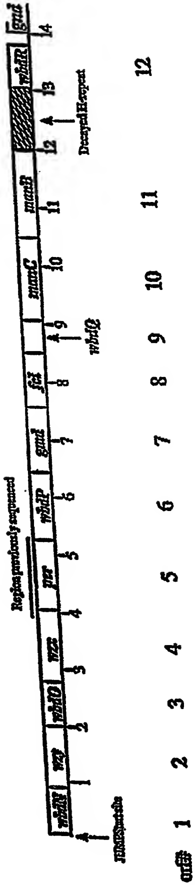
gnd
wbdR
manB
manC
wbdQ
fnl
gmd
wbdP
Region previously sequenced
per
wzx
wbdO
wzt
1
2
3
4
5
6
7
8
9
10
11
12

Figure 4

rmlB
rmlD
rmlA
rmlC
ddhD
ddhA
ddhB
ddhC
abe
wzx
whaR
whaL
whaQ
wzy
whaW
whaZ
manC
manB
whaP
0
5
10
kb

Figure 5

rml
B
D
A
C
D
A
B
C
ddh
abe
wzx
wbaV
wbaU
wbaN
manC
manB
whaP
0
5
10
kb

Figure 6

```
GATCTGATGGCCGTAGGGCGCTACGTGCTTTCTGCTGATATCTGGGCTGAGTTGGAAAAA      60

ACTGCTCCAGGTGCCTGGGGACGTATTCAACTGACTGATGCTATTGCAGAGTTGGCTAAA     120

AAACAGTCTGTTGATGCCATGCTGATGACCGGCGACAGCTACGACTGCGGTAAGAAGATG     180

GGCTATATGCAGGCATTCGTTAAGTATGGGCTGCGCAACCTTAAAGAAGGGGCGAAGTTC     240

CGTAAGAGCATCAAGAAGCTACTGAGTGAGTAGAGATTTACACGTCTTTGTGACGATAAG     300

CCAGAAAAAATAGCGGCAGTTAACATCCAGGCTTCTATGCTTTAAGCAATGGAATGTTAC     360

TGCCGTTTTTTATGAAAAATGACCAATAATAACAAGTTAACCTACCAAGTTTAATCTGCT     420

TTTTGTTGGATTTTTTCTTGTTTCTGGTCGCATTTGGTAAGACAATTAGCGTGAGTTTTA     480

GAGAGTTTTGCGGGATCTCGCGGAACTGCTCACATCTTTGGCATTTAGTTAGTGCACTGG     540

TAGCTGTTAAGCCAGGGGCGGTAGCTTGCCTAATTAATTTTTAACGTATACATTTATTCT     600

TGCCGCTTATAGCAAATAAAGTCAATCGGATTAAACTTCTTTTCCATTAGGTAAAAGAGT     660

GTTTGTAGTCGCTCAGGGAAATTGGTTTTGGTAGTAGTACTTTTCAAATTATCCATTTTC     720

                    Start of orf1
                  M  L  L  C  C  I  H  I  N  V  Y  Y  L  L
CGATTTAGATGGCAGTTGATGTTACTATGCTGCATACATATCAATGTATATTATTTACTT     780

 L  E  C  D  M  K  K  I  V  I  I  G  N  V  A  S  M  M  L  R
TTAGAATGTGATATGAAAAAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGG     840

 F  R  K  E  L  I  M  N  L  V  R  Q  G  D  N  V  Y  C  L  A
TTCAGGAAAGAATTAATCATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCA     900

 N  D  F  S  T  E  D  L  K  V  L  S  S  W  G  V  K  G  V  K
AATGATTTTTCCACTGAAGATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAA     960

 F  S  L  N  S  K  G  I  N  P  F  K  D  I  I  A  V  Y  E  L
TTCTCTCTTAACTCAAAGGGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTA    1020

 K  K  I  L  K  D  I  S  P  D  I  V  F  S  Y  F  V  K  P  V
AAAAAAATTCTTAAGGATATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTA    1080

 I  F  G  T  I  A  S  K  L  S  K  V  P  R  I  V  G  M  I  E
ATATTTGGAACTATTGCTTCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAA    1140

 G  L  G  N  A  F  T  Y  Y  K  G  K  Q  T  T  K  T  K  M  I
GGTCTAGGTAATGCCTTCACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATA    1200

 K  W  I  Q  I  L  L  Y  K  L  A  L  P  M  L  D  D  L  I  L
AAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTA    1260

 L  N  H  D  D  K  K  D  L  I  D  Q  Y  N  I  K  A  K  V  T
TTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACA    1320

 V  L  G  G  I  G  L  D  L  N  E  F  S  Y  K  E  P  P  K  E
GTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAG    1380

 K  I  T  F  I  F  I  A  R  L  L  R  E  K  G  I  F  E  F  I
AAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATT    1440

 E  A  A  K  F  V  K  T  T  Y  P  S  S  E  F  V  I  L  G  G
GAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGT    1500
```

Figure 7/1

```
 F  E  S  N  N  P  F  S  L  Q  K  N  E  I  E  S  L  R  K  E
TTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAA    1560

 H  D  L  I  Y  P  G  H  V  E  N  V  Q  D  W  L  E  K  S  S
CATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCT    1620

 V  F  V  L  P  T  S  Y  R  E  G  V  P  R  V  I  Q  E  A  M
GTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATG    1680

 A  I  G  R  P  V  I  T  T  N  V  P  G  C  R  D  I  I  N  D
GCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGAT    1740

 G  V  N  G  F  L  I  P  P  F  E  I  N  L  L  A  E  K  M  K
GGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAA    1800

 Y  F  I  E  N  K  D  K  V  L  E  M  G  L  A  G  R  K  F  A
TATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCA    1860

 E  K  N  F  D  A  F  E  K  N  N  R  L  A  S  I  I  K  S  N
GAAAAAAACTTTGATGCTTTTGAAAAAAATAATAGACTAGCATCAATAATAAAATCAAAT    1920

                End of orf1
 N  D  F  *
AATGATTTTTGACTTGAGCAGAAATTATTTATATTTCAATCTGAAAAATAAAGGCTGTTA    1980

   Start of orf2
   M  N  K  V  A  L  I  T  G  I  T  G  Q  D  G  S  Y  L  A
TTATGAATAAAGTGGCATTAATTACTGGTATCACTGGGCAAGATGGCTCCTATTTGGCAG    2040

 E  L  L  L  E  K  G  Y  E  V  H  G  I  K  R  R  A  S  S  F
AATTATTGTTAGAAAAAGGTTATGAAGTTCATGGTATTAAACGCCGTGCATCTTCATTTA    2100

 N  T  E  R  V  D  H  I  Y  Q  D  S  H  L  A  N  P  K  L  F
ATACTGAGCGAGTGGATCACATCTATCAGGATTCACATTTAGCTAATCCTAAACTTTTTC    2160

 L  H  Y  G  D  L  T  D  T  S  N  L  T  R  I  L  K  E  V  Q
TACACTATGGCGATTTGACAGATACTTCCAATCTGACCCGTATTTTAAAAGAAGTTCAAC    2220

 P  D  E  V  Y  N  L  G  A  M  S  H  V  A  V  S  F  E  S  P
CAGATGAAGTTTACAATTTGGGGGCGATGAGCCATGTAGCGGTATCATTTGAGTCACCAG    2280

 E  Y  T  A  D  V  D  A  I  G  T  L  R  L  L  E  A  I  R  I
AATACACTGCTGATGTTGATGCGATAGGAACATTGCGTCTTCTTGAAGCTATCAGGATAT    2340

 L  G  L  E  K  K  T  K  F  Y  Q  A  S  T  S  E  L  Y  G  L
TGGGGCTGGAAAAAAAGACAAAATTTTATCAGGCTTCAACTTCAGAGCTTTATGGTTTGG    2400

 V  Q  E  I  P  Q  K  E  T  T  P  F  Y  P  R  S  P  Y  A  V
TTCAAGAAATTCCACAAAAAGAGACTACGCCATTTTATCCACGTTCGCCTTATGCTGTTG    2460

 A  K  L  Y  A  Y  W  I  T  V  N  Y  R  E  S  Y  G  M  F  A
CAAAATTATATGCCTATTGGATCACTGTTAATTATCGTGAGTCTTATGGTATGTTTGCCT    2520

 C  N  G  I  L  F  N  H  E  S  P  R  R  G  E  T  F  V  T  R
GCAATGGTATTCTCTTTAACCACGAATCACCTCGCCGTGGCGAGACCTTTGTTACTCGTA    2580

 K  I  T  R  G  I  A  N  I  A  Q  G  L  D  K  C  L  Y  L  G
AAATAACACGCGGGATAGCAAATATTGCTCAAGGTCTTGATAAATGCTTATACTTGGGAA    2640

 N  M  D  S  L  R  D  W  G  H  A  K  D  Y  V  K  M  Q  W  M
ATATGGATTCTCTGCGTGATTGGGGACATGCTAAGGATTATGTCAAAATGCAATGGATGA    2700
```

Figure 7/2

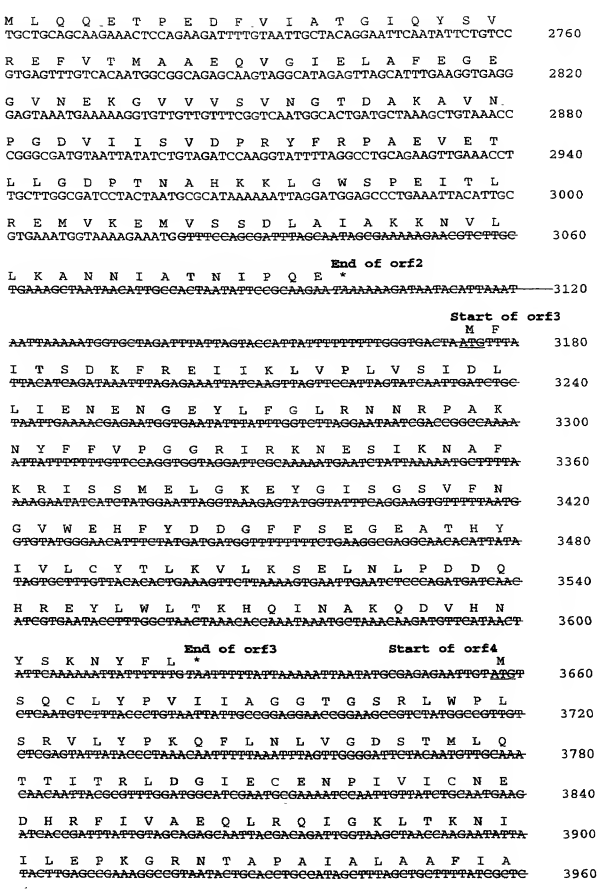
M L Q Q E T P E D F V I A T G I Q Y S V
TGCTGCAGCAAGAAACTCCAGAAGATTTTGTAATTGCTACAGGAATTCAATATTCTGTCC 2760
R E F V T M A A E Q V G I E L A F E G E
GTGAGTTTGTCACAATGGCGGCAGAGCAAGTAGGCATAGAGTTAGCATTTGAAGGTGAGG 2820
G V N E K G V V V S V N G T D A K A V N
GAGTAAATGAAAAAGGTGTTGTTGTTTCGGTCAATGGCACTGATGCTAAAGCTGTAAACC 2880
P G D V I I S V D P R Y F R P A E V E T
CGGGCGATGTAATTATATCTGTAGATCCAAGGTATTTTAGGCCTGCAGAAGTTGAAACCT 2940
L L G D P T N A H K K L G W S P E I T L
TGCTTGGCGATCCTACTAATGCGCATAAAAAATTAGGATGGAGCCCTGAAATTACATTGC 3000
R E M V K E M V S S D L A I A K K N V L
GTGAAATGGTAAAAGAAATGGTTTCCAGCGATTTAGCAATAGCGAAAAAGAACGTCTTGC 3060
End of orf2
L K A N N I A T N I P Q E *
TGAAAGCTAATAACATTGCCACTAATATTCCGCAAGAATAAAAAAGATAATACATTAAAT 3120
Start of orf3
M F
AATTAAAAATGGTGCTAGATTTATTAGTACCATTATTTTTTTTGGGTGACTAATGTTTA 3180
I T S D K F R E I I K L V P L V S I D L
TTACATCAGATAAATTTAGAGAAATTATCAAGTTAGTTCCATTAGTATCAATTGATCTGC 3240
L I E N E N G E Y L F G L R N N R P A K
TAATTGAAAACGAGAATGGTGAATATTTATTTGGTCTTAGGAATAATCGACCGGCCAAAA 3300
N Y F F V P G G R I R K N E S I K N A F
ATTAATTTTTTTGTTCCAGGTGGTAGGATTCGCAAAAATGAATCTATTAAAAATGCTTTTA 3360
K R I S S M E L G K E Y G I S G S V F N
AAAGAATATCATCTATGGAATTAGGTAAAGAGTATGGTATTTCAGGAAGTGTTTTTAATG 3420
G V W E H F Y D D G F F S E G E A T H Y
GTGTATGGGAACATTTCTATGATGATGGTTTTTTTTCTGAAGGCGAGGCAACACATTATA 3480
I V L C Y T L K V L K S E L N L P D D Q
TAGTGCTTTGTTACACACTGAAAGTTCTTAAAAGTGAATTGAATCTCCCAGATGATCAAC 3540
H R E Y L W L T K H Q I N A K Q D V H N
ATCGTGAATACCTTTGGCTAACTAAACACGAAATAAATGCTAAACAAGATGTTCATAACT 3600
End of orf3
Start of orf4
Y S K N Y F L * M
ATTCAAAAAATTATTTTTTGTAATTTTTATTAAAAATTAATATGCGAGAGAATTGTATGT 3660
S Q C L Y P V I I A G G T G S R L W P L
CTCAATGTCTTTACCCTGTAATTATTGCCGGAGGAACGGAAGCCGTCTATGGCCGTTGT 3720
S R V L Y P K Q F L N L V G D S T M L Q
CTCGAGTATTATACCCTAAACAATTTTTAATTTAGTTGGGGATTCTACAATGTTGCAAA 3780
T T I T R L D G I E C E N P I V I C N E
CAACAATTACGCGTTTGGATGGCATCGAATGCGAAAATCCAATTGTTATCTGCAATGAAG 3840
D H R F I V A E Q L R Q I G K L T K N I
ATCACCGATTTATTGTAGCAGAGCAATTACGACAGATTGGTAAGCTAACAAAGAATATTA 3900
I L E P K G R N T A P A I A L A A F I A
TACTTGAGCCGAAAGGCGTAATACTGCACCTGCCATAGCTTAGCTGCTTTATCGCTC 3960

Figure 7/3

```
Q K N N P N D D P L L L V L A A D H S I
AGAAGAATAATCCTAATGACGACCCTTTATTATTAGTACTTGCGGCAGACCACTCTATAA 4020

N N E K A F R E S I I K A M P Y A T S G
ATAATGAAAAAGCATTTCGAGAGTCAATAATAAAAGCTATGCCGTATGCAACTTCTGGGA 4080

K L V T F G I I P D T A N T G Y G Y I K
AGTTAGTAACATTTGGAATTATTCCGGACACGGCAAATACTGGTTATGGATATATTTAAGA 4140

R S S S A D P N K E F P A Y N V A E F V
GAAGTTCTTCAGCTGATCCTAATAAAGAATTCCCAGCATATAATGTTGCGGAGTTTGTAG 4200

E K P D V K T A Q E Y I S S G N Y Y W N
AAAAACCAGATGTTAAAACAGCACAGGAATATATTTCGAGTGGGAATTATTACTGGAATA 4260

S G M F L F R A S K Y L D E L R K F R P
GCGGAATGTTTTTTATTTCGCGCCAGTAAATATCTTGATGAACTACGGAAATTTAGACCAG 4320

D I Y H S C E C A T A T A N I D M D F V
ATATTTATCATAGCTGTGAATGTGCAACCGCTACAGCAAATATAGATATGGACTTTGTCC 4380

R I N E A E F I N C P E E S I D Y A V M
GAATTAACGAGGCTGAGTTTATTAATTGTCCTGAAGAGTCTATCGATTATGCTGTGATGG 4440

E K T K D A V V L P I D I G W N D V G S
AAAAACAAAAGACGCTGTAGTTCTTCCGATAGATATTGGCTGGAATGACGTGGGTTCTT 4500

W S S L W D I S Q K D C H G N V C H G D
GGTCATCACTTTGGGATATAAGCCAAAAGGATTGCCATGGTAATGTGTGCCATGGGGATG 4560

V L N H D G E N S F I Y S E S S L V A T
TGCTCAATCATGATGGAGAAAATAGTTTTATTTACTCTGAGTCAAGTCTGGTTGCGACAG 4620

V G V S N L V I V Q T K D A V L V A D R
TCGGAGTAAGTAATTTAGTAATTGTCCAAACCAAGGATGCTGTACTGGTTGCGGACCGTG 4680

D K V Q N V K N I V D D L K K R K R A E
ATAAAGTCCAAAATGTTAAAAACATAGTTGACGATCTAAAAAAGAAAAAGCGTGCTGAAT 4740

Y Y M H R A V F R P W G K F D A I D Q G
ACTACATGCATCGTGCAGTTTTCGCCCTTGGGGTAAATTCGATGCAATAGACCAAGGCG 4800

D R Y R V K K I I V K P G E G L D L R M
ATAGATATAGAGTAAAAAAATAATAGTTAAACCAGGAGAAGGCTTAGATTTAAGGATGC 4860

H H H R A E H W I V V S G T A K V S L G
ATCATCATAGGGCAGAGCATTGGATTGTTGTATCCGGTACTGCTAAAGTTTCACTAGGTA 4920

S E V K L L V S N E S I Y I P Q G A K Y
GTGAAGTTAAACTAATTAGTTTCTAATGAGTCTATATATATCCCTCAGGGAGCAAAATATA 4980

S L E N P G V I P L H L I E V S S G D Y
GTCTTGAGAATCCAGGCGTAATACCTTTGCATCTAATTGAAGTAAGTTCTGGTGATTACC 5040

L E S D D I V R F T D R Y N S K Q F L K
TTGAATCAGATGATATAGTGCGTTTTACTGACAGATATAACAGTAAACAATTCCTAAAGC 5100
```

End of orf4 Start of orf5

```
M N K I T C F K A Y D I R G R L
R D *
GAGATTGATAAATATGAATAAAATTAACTTGCTTCAAAGCATATGATATACGTGGGCGTCT 5160
```

Figure 7/4

```
G A E L N D E I A Y R I G R A Y G E F F
TGGTGCTGAATTGAATGATGAAATAGCATATAGAATTGGTCGCGCTTATGGTGAGTTTTT  5220

K P Q T V V V G G D A R L T S E S L K K
TAAACCTCAAACTGTAGTTGTGGGAGGAGATGCTCGCTTAACAAGTGAGAGTTTAAAGAA  5280

S L S N G L C D A G V N V L D L G M C G
ATCACTCTCAAATGGGCTATGTGATGCAGGCGTAAATGTCTTAGATCTTGGAATGTGTGG  5340

T E E I Y F S T W Y L G I D G G I E V T
TACTGAAGAGATATATTTTTCCACTTGGTATTTAGGAATTGATGGTGGAATCGAGGTAAC  5400

A S H N P I D Y N G M K L V T K G A R P
TGCAAGCCATAATCCAATTGATTATAATGGAATGAAATTAGTAACCAAAGGTGCTCGACC  5460

I S S D T G L K D I Q Q L V E S N N F E
AATCAGCAGTGACACAGGTCTCAAAGATATACAACAATTAGTAGAGAGTAATAATTTTGA  5520

E L N L E K K G N I T K Y S T R D A Y I
AGAGCTCAACCTAGAAAAAAAAGGGAATATTACCAAATATTCCACCCGAGATGCCTACAT  5580

N H L M G Y A N L Q K I K K I K I V V N
AAATCATTTGATGGGCTATGCTAATCTGCAAAAAATAAAAAAAATCAAAATAGTTGTGAA  5640

S G N G A A G P V I D A I E E C F L R N
TTCTGGGAATGGTGCAGCTGGTCCTGTTATTGATGCTATTGAGGAATGCTTTTTACGGAA  5700

N I P I Q F V K I N N T P D G N F P H G
CAATATTCCGATTCAGTTTGTAAAAATAAATAATACACCCGATGGTAATTTTCCACATGG  5760

I P N P L L P E C R E D T S S A V I R H
TATCCCTAATCCATTACTACCTGAGTGCAGAGAAGATACCAGCAGTGCGGTTATAAGACA  5820

S A D F G I A F D G D F D R C F F F D E
TAGTGCTGATTTGGTATTGCATTTGATGGTGATTTTGATAGGTGTTTTTTCTTTGATGA  5880

N G Q F I E G Y Y I V G L L A E V F L G
AAATGGACAATTTATTGAAGGATACTACATTGTTGGTTTATTAGCGGAAGTTTTTTTAGG  5940

K Y P N A K I I H D P R L I W N T I D I
GAAATATCCAAACGCAAAAATCATTCATGATCCTGCCCTTATATGGAATACTATTGATAT  6000

V E S H G G I P I M T K T G H A Y I K Q
CGTAGAAAGTCATGGTGGTATACCTATAATGACTAAAACCGGTCATGCTTACATTAAGCA  6060

R M R E E D A V Y G G E M S A H H Y F K
AAGAATGCGTGAAGAGGATGCCGTATATGGCGGCGAAATGAGTGCGCATCATTATTTTAA  6120

D F A Y C D S G M I P W I L I C E L L S
AGATTTTGCATACTGCGATAGTGGAATGATTCCTTGGATTTAATTTGTGAACTTTTGAG  6180

L T N K K L G E L V C G C I N D W P A S
TCTGACAAATAAAAAATTAGGTGAACTGGTTTGTGGTTGTATAAACGACTGGCCGGCAAG  6240

G E I N C T L D N P Q N E I D K L F N R
TGGAGAAATAACTGTACACTAGACAATCCGCAAAATGAAATAGATAAATTATTTAATCG  6300

Y K D S A L A V D Y T D G L T M E F S D
TTACAAAGATAGTGCCTTAGCTGTTGATTACACTGATGGATTAACTATGGAGTTCTCTGA  6360

W R F N V R C S N T E P V V R L N V E S
TTGGCGTTTTAATGTTAGATGCTCAAATACAGAACCTGTAGTACGATTGAATGTAGAATC  6420

R N N A I L M Q E K T E E I L N F I S K
TAGGAATAATGCTATTCTTATGCAGGAAAAACAGAAGAAATTCTGAATTTTATATTCAAA  6480
```

Figure 7/5

End of orf5 Start of orf6

```
                                                  M  K  V  L  L  T  G
ATAAATTTGCACCTGAGTTCATAATGGGAACAAGAAATATATGAAAGTACTTCTGACTGG      6540

 S  T  G  M  V  G  K  N  I  L  E  H  D  S  A  S  K  Y  N  I
CTCAACTGGCATGGTTGGTAAGAATATATTAGAGCATGATAGTGCAAGTAAATATAATAT      6600

 L  T  P  T  S  S  D  L  N  L  L  D  K  N  E  I  E  K  F  M
ACTTACTCCAACCAGCTCTGATTTGAATTTATTAGATAAAAATGAAATAGAAAAATTCAT      6660

 L  I  N  M  P  D  C  I  I  H  A  A  G  L  V  G  G  I  H  A
GCTTATCAACATGCCAGACTGTATTATACATGCAGCGGGATTAGTTGGAGGCATTCATGC      6720

 N  I  S  R  P  F  D  F  L  E  K  N  L  Q  M  G  L  N  L  V
AAATATAAGCAGGCCGTTTGATTTTCTGGAAAAAAATTTGCAGATGGGTTTAAATTTAGT      6780

 S  V  A  K  K  L  G  I  K  K  V  L  N  L  G  S  S  C  M  Y
TTCCGTCGCAAAAAAACTAGGTATCAAGAAAGTGCTTAACTTGGGTAGTTCATGCATGTA      6840

 P  K  N  F  E  E  A  I  P  E  K  A  L  L  T  G  E  L  E  E
CCCCAAAAACTTTGAAGAGGCTATTCCTGAGAAAGCTCTGTTAACTGGTGAGCTAGAAGA      6900

 T  N  E  G  Y  A  I  A  K  I  A  V  A  K  A  C  E  Y  I  S
AACTAATGAGGGATATGCTATTGCGAAAATTGCTGTAGCAAAAGCATGCGAATATATATC      6960

 R  E  N  S  N  Y  F  Y  K  T  I  I  P  C  N  L  Y  G  K  Y
AAGAGAAAACTCTAATTATTTTTATAAAACAATTATCCCATGTAATTTATATGGGAAATA      7020

 D  K  F  D  D  N  S  S  H  M  I  P  A  V  I  K  K  I  H  H
TGATAAATTTGATGATAACTCGTCACATATGATTCCGGCAGTTATAAAAAAAATCCATCA      7080

 A  K  I  N  N  V  P  E  I  E  I  W  G  D  G  N  S  R  R  E
TGCGAAAATTAATAATGTCCCAGAGATCGAAATTTGGGGGGATGGTAATTCGCGCCGTGA      7140

 F  M  Y  A  E  D  L  A  D  L  I  F  Y  V  I  P  K  I  E  F
GTTTATGTATGCAGAAGATTTAGCTGATCTTATTTTTTATGTTATTCCTAAAATAGAATT      7200

 M  P  N  M  V  N  A  G  L  G  Y  D  Y  S  I  N  D  Y  Y  K
CATGCCTAATATGGTAAATGCTGGTTTAGGTTACGATTATTCAATTAATGACTATTATAA      7260

 I  I  A  E  E  I  G  Y  T  G  S  F  S  H  D  L  T  K  P  T
GATAATTGCAGAAGAAATTGGTTATACTGGGAGTTTTTCTCATGATTTAACAAAACCAAC      7320

 G  M  K  R  K  L  V  D  I  S  L  L  N  K  I  G  W  S  S  H
AGGAATGAAACGGAAGCTAGTAGATATTTCATTGCTTAATAAAATTGGTTGGTCAAGTCA      7380

 F  E  L  R  D  G  I  R  K  T  Y  N  Y  Y  L  E  N  Q  N  K
CTTTGAACTCAGAGATGGCATCAGAAAGACCTATAATTATTACTTGGAGAATCAAAATAA      7440
```

Start of orf7, End of orf6

```
 M  I  T  Y  P  L  A  S  N  T  W  D  E  Y  E  Y  A  A  I  Q
ATGATTACATACCCACTTGCTAGTAATACTTGGGATGAATATGAGTATGCAGCAATACAG      7500

 S  V  I  D  S  K  M  F  T  M  G  K  K  V  E  L  Y  E  K  N
TCAGTAATTGACTCAAAAATGTTTACCATGGGTAAAAAGGTTGAGTTATATGAGAAAAAT      7560

 F  A  D  L  F  G  S  K  Y  A  V  M  V  S  S  G  S  T  A  N
TTTGCTGATTTGTTTGGTAGCAAATATGCCGTAATGGTTAGCTCTGGTTCTACAGCTAAT      7620
```


Figure 7/6

```
L  L  M  I  A  A  L  F  F  T  N  K  P  K  L  K  R  G  D  E
CTGTTAATGATTGCTGCCCTTTTCTTCACTAATAAACCAAAACTTAAAAGAGGTGATGAA  7680

I  I  V  P  A  V  S  W  S  T  T  Y  Y  P  L  Q  Q  Y  G  L
ATAATAGTACCTGCAGTGTCATGGTCTACGACATATTACCCTCTGCAACAGTATGGCTTA  7740

K  V  K  F  V  D  I  N  K  E  T  L  N  I  D  I  D  S  L  K
AAGGTGAAGTTTGTCGATATCAATAAAGAAACTTTAAATATTGATATCGATAGTTTGAAA  7800

N  A  I  S  D  K  T  K  A  I  L  T  V  N  L  L  G  N  P  N
AATGCTATTTCAGATAAAACAAAAGCAATATTGACAGTAAATTTATTAGGTAATCCTAAT  7860

D  F  A  K  I  N  E  I  I  N  N  R  D  I  I  L  L  E  D  N
GATTTTGCAAAAATAAATGAGATAATAAATAATAGGGATATTATCTTACTAGAAGATAAC  7920

C  E  S  M  G  A  V  F  Q  N  K  Q  A  G  T  F  G  V  M  G
TGTGAGTCGATGGGCGCGGTCTTTCAAAATAAGCAGGCAGGCACATTCGGAGTTATGGGT  7980

T  F  S  S  F  Y  S  H  H  I  A  T  M  E  G  G  C  V  V  T
ACCTTTAGTTCTTTTTACTCTCATCATATAGCTACAATGGAAGGGGGCTGCGTAGTTACT  8040

D  D  E  E  L  Y  H  V  L  L  C  L  R  A  H  G  W  T  R  N
GATGATGAAGAGCTGTATCATGTATTGTTGTGCCTTCGAGCTCATGGTTGGACAAGAAAT  8100

L  P  K  E  N  M  V  T  G  T  K  S  D  D  I  F  E  E  S  F
TTACCAAAAGAGAATATGGTTACAGGCACTAAGAGTGATGATATTTTCGAAGAGTCGTTT  8160

K  F  V  L  P  G  Y  N  V  R  P  L  E  M  S  G  A  I  G  I
AAGTTTGTTTTACCAGGATACAATGTTCGCCCACTTGAAATGAGTGGTGCTATTGGGATA  8220

E  Q  L  K  K  L  P  G  F  I  S  T  R  R  S  N  A  Q  Y  F
GAGCAACTTAAAAAGTTACCAGGTTTTATATCCACAGACGTTCCAATGCACAATATTTT  8280

V  D  K  F  K  D  H  P  F  L  D  I  Q  K  E  V  G  E  S  S
GTAGATAAATTTAAAGATCATCCATTCCTTGATATACAAAAAGAAGTTGGTGAAAGTAGC  8340

W  F  G  F  S  F  V  I  K  E  G  A  A  I  E  R  K  S  L  V
TGGTTTGGTTTTTCCTTCGTTATAAGGAGGGAGCTGCTATTGAGAGGAAGAGTTTAGTA  8400

N  N  L  I  S  A  G  I  E  C  R  P  I  V  T  G  N  F  L  K
AATAATCTGATCTCAGCAGGCATTGAATGCCGACCAATTGTTACTGGGAATTTTCTCAAA  8460

N  E  R  V  L  S  Y  F  D  Y  S  V  H  D  T  V  A  N  A  E
AATGAACGTGTTTTGAGTTATTTTGATTACTCTGTACATGATACGGTAGCAAATGCCGAA  8520

Y  I  D  K  N  G  F  F  V  G  N  H  Q  I  P  L  F  N  E  I
TATATAGATAAGAATGGTTTTTTTTGTCGGAAACCACCAGATACCTTTGTTTAATGAAATA  8580

                    End of orf7
D  Y  L  R  K  V  L  K  *
GATTATCTACGAAAAGTATTAAAATAACTAACGAGGCACTCTATTTCGAATAGAGTGCCT  8640

          Start of orf8
          M  V  L  T  V  K  K  I  L  A  F  G  Y  S  K  V  L  P
TTAAGATGGTATTAACAGTGAAAAAAATTTAGCGTTTGGCTATTCTAAAGTACTACCAC  8700

P  V  I  E  Q  F  V  N  P  I  C  I  F  I  I  T  P  L  I  L
CGGTTATTGAACAGTTTGTCAATCCAATTTGCATCTTCATTATCACACCACTAATACTCA  8760

N  H  L  G  K  Q  S  Y  G  N  W  I  L  L  I  T  I  V  S  F
ACCACCTGGGTAAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTT  8820
```

Figure 7/7

```
S  Q  L  I  C  G  G  C  S  A  W  I  A  K  I  I  A  E  Q  R
CTCAGTTAATATGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAA  8880

I  L  S  D  L  S  K  K  N  A  L  R  Q  I  S  Y  N  F  S  I
TTCTTAGTGATTTATCAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTC  8940

V  I  I  A  F  A  V  L  I  S  F  L  I  L  S  I  C  F  F  D
TTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATG  9000

V  A  R  N  N  S  S  F  L  F  A  I  I  I  C  G  F  F  Q  E
TTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTCAGGAAG  9060

V  D  N  L  F  S  G  A  L  K  G  F  E  K  F  N  V  S  C  F
TTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTT  9120

F  E  V  I  T  R  V  L  W  A  S  I  V  I  Y  G  I  Y  G  N
TTGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATG  9180

A  L  L  Y  F  T  C  L  A  F  T  I  K  G  M  L  K  Y  I  L
CACTCTTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTG  9240

V  C  L  N  I  T  G  C  F  I  N  P  N  F  N  R  V  G  I  V
TATGTCTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTA  9300

N  L  L  N  E  S  K  W  M  F  L  Q  L  T  G  G  V  S  L  S
ATTTGTTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTT  9360

L  F  D  R  L  V  I  P  L  I  L  S  V  S  K  L  A  S  Y  V
TGTTTGATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCC  9420

P  C  L  Q  L  A  Q  L  M  F  T  L  S  A  S  A  N  Q  I  L
CTTGCCTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTAC  9480

L  P  M  F  A  R  M  K  A  S  N  T  F  P  S  N  C  F  F  K
TACCAATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAA  9540

I  L  L  V  S  L  I  S  V  L  P  C  L  A  L  F  F  F  G  R
TTCTGCTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTC  9600

D  I  L  S  I  W  I  N  P  T  F  A  T  E  N  Y  K  L  M  Q
ATATATTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAA  9660

I  L  A  I  S  Y  I  L  L  S  M  M  T  S  F  H  F  L  L
TTTTAGCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAG  9720

G  I  G  K  S  K  L  V  A  N  L  N  L  V  A  G  L  A  L  A
GAATTGGTAAAATCTAAGCTTGTTGCAAATTTAAATCTGGTTGCAGGCCTCGCACTTGCTG  9780

A  S  T  L  I  A  A  H  Y  G  L  Y  A  I  S  M  V  K  I  I
CTTCAACGTTAATCGCAGCTCATTATGGCCTTTATGCAATATCTATGGTAAAAATAATAT  9840

Y  P  A  F  Q  F  Y  Y  L  Y  V  A  F  V  Y  F  N  R  A  K
ATCCGGCTTTTCAATTTATTACCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAA  9900
```

Start of orf9, End of orf8

```
   M  S  I  D  L  L  F  S  I  T  E  I  A  I  V  F  S  C  T  I
N  V  Y  *
ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT  9960

   Y  I  F  T  Q  C  L  L  M  R  R  I  Y  L  D  K  S  I  L  I
TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT  10020

   L  L  C  L  L  F  F  L  V  I  I  Q  L  P  E  L  N  V  N  G
CTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT  10080
```

Figure 7/8

```
  L  V  D  S  L  K  L  S  L  P  L  L  M  V  F  I  A  F  Q  K
TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA   10140

  P  K  L  C  L  W  V  I  I  A  L  L  F  L  N  S  A  F  N  F
CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT   10200

  L  Y  L  K  T  F  D  K  F  S  S  F  P  F  T  F  F  I  L  L
TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTG   10260

  F  Y  L  F  R  L  G  I  G  N  L  P  V  Y  K  N  K  K  F  Y
TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC   10320

  A  L  I  F  L  F  I  L  I  D  I  M  Q  S  L  L  I  N  Y  R
GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG   10380

  G  Q  I  L  Y  S  V  I  C  I  L  I  L  V  F  K  V  N  L  R
GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA   10440

  K  K  I  P  Y  F  F  L  M  L  P  V  L  Y  V  I  I  M  A  Y
AAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTAT   10500

  I  G  F  N  Y  F  N  K  G  V  T  F  F  E  P  T  A  S  N  I
ATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATT   10560

  E  R  T  G  M  I  Y  Y  L  V  S  Q  L  G  D  Y  I  F  H  G
GAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGT   10620

  M  G  T  L  N  F  L  N  N  G  G  Q  Y  K  T  L  Y  G  L  P
ATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCA   10680

  S  L  I  P  N  D  P  H  D  F  L  L  R  F  F  I  S  I  G  V
TCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTG   10740

  I  G  A  L  V  Y  H  S  I  F  F  V  F  F  R  R  I  S  F  L
ATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTA   10800

  L  Y  E  R  N  A  P  F  I  V  V  S  C  L  L  L  L  Q  V  V
TTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTG   10860

  L  I  Y  T  L  N  P  F  D  A  F  N  R  L  I  C  G  L  T  V
TTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTT   10920
```

Start of orf10 **End of orf9**

```
  G  V  V  Y  G  F  A  K  I  R  *
        M  D  L  Q  K  L  D  K  Y  T  C  N  G  N  L  D  A
GGAGTTGTTTATGGATTTGCAAAAATTAGATAAGTATACCTGTAATGGAAATTTAGACGC   10980

  P  L  V  S  I  I  I  A  T  Y  N  S  E  L  D  I  A  K  C  L
TCCACTTGTTTCAATAATCATTGCAACTTATAATTCTGAACTTGATATAGCTAAGTGTTT   11040

  Q  S  V  T  N  Q  S  Y  K  N  I  E  I  I  I  M  D  G  G  S
GCAATCGGTAACTAATCAATCTTATAAGAATATTGAAATCATAATAATGGATGGAGGATC   11100

  S  D  K  T  L  D  I  A  K  S  F  K  D  D  R  I  K  I  V  S
TTCTGATAAAACGCTTGATATTGCAAAATCGTTTAAAGACGACCGAATAAAAATAGTTTC   11160

  E  K  D  R  G  I  Y  D  A  W  N  K  A  V  D  L  S  I  G  D
AGAGAAAGATCGTGGAATTTATGATGCCTGGAATAAAGCAGTTGATTTATCCATTGGTGA   11220

  W  V  A  F  I  G  S  D  D  V  Y  Y  H  T  D  A  I  A  S  L
TTGGGTAGCATTTATTGGTTCAGATGATGTTTACTATCATACAGATGCAATTGCTTCATT   11280

  M  K  G  V  M  V  S  N  G  A  P  V  V  Y  G  R  T  A  H  E
GATGAAGGGGGTTATGGTATCTAATGGCGCCCCTGTGGTTTATGGGAGGACAGCGCACGA   11340
```

Figure 7/9

```
  G  P  D  R  N  I  S  G  F  S  G  S  E  W  Y  N  L  T  G  F
AGGTCCCGATAGGAACATATCTGGATTTTCAGGCAGTGAATGGTACAACCTAACAGGATT  11400

  K  F  N  Y  Y  K  C  N  L  P  L  P  I  M  S  A  I  Y  S  R
TAAGTTTAATTATTACAAATGTAATTTACCATTGCCCATTATGAGCGCAATATATTCTCG  11460

  D  F  F  R  N  E  R  F  D  I  K  L  K  I  V  A  D  A  D  W
TGATTTCTTCAGAAACGAACGTTTTGATATTAAATTAAAAATTGTTGCTGACGCTGATTG  11520

  F  L  R  C  F  I  K  W  S  K  E  K  S  P  Y  F  I  N  D  T
GTTTCTGAGATGTTTCATCAAATGGAGTAAAGAGAAGTCACCTTATTTTATTAATGACAC  11580

  T  P  I  V  R  M  G  Y  G  G  V  S  T  D  I  S  S  Q  V  K
GACCCCTATTGTTAGAATGGGATATGGTGGGGTTTCGACTGATATTTCTTCTCAAGTTAA  11640

  T  T  L  E  S  F  I  V  R  K  K  N  N  I  S  C  L  N  I  Q
AACTACGCTAGAAAGTTTCATTGTACGCAAAAAGAATAATATATCCTGTTTAAACATACA  11700

  L  I  L  R  Y  A  K  I  L  V  M  V  A  I  K  N  I  F  G  N
GCTGATTCTTAGATATGCTAAAATTCTGGTGATGGTAGCGATCAAAAATATTTTTGGCAA  11760

  N  V  Y  K  L  M  H  N  G  Y  H  S  L  K  K  I  K  N  K  I
TAATGTTTATAAATTAATGCATAACGGGTATCATTCCCTAAAGAAAATCAAGAATAAAAT  11820

Start of orf11, End of orf10
  M  K  I  V  Y  I  I  T  G  L  T  C  G  G  A  E  H  L  M  T
  *
ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG  11880

  Q  L  A  D  Q  M  F  I  R  G  H  D  V  N  I  I  C  L  T  G
CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT  11940

  I  S  E  V  K  P  T  Q  N  I  N  I  H  Y  V  N  M  D  K  N
ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT  12000

  F  R  S  F  F  R  A  L  F  Q  V  K  K  I  I  V  A  L  K  P
TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA  12060

  D  I  I  H  S  H  M  F  H  A  N  I  F  S  R  F  I  R  M  L
GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG  12120

  I  P  A  V  P  L  I  C  T  A  H  N  K  N  E  G  G  N  A  R
ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG  12180

  M  F  C  Y  R  L  S  D  F  L  A  S  I  T  T  N  V  S  K  E
ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG  12240

  A  V  Q  E  F  I  A  R  K  A  T  P  K  N  K  I  V  E  I  P
GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG  12300

  N  F  I  N  T  N  K  F  D  F  D  I  N  V  R  K  K  T  R  D
AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT  12360

  A  F  N  L  K  D  S  T  A  V  L  L  A  V  G  R  L  V  E  A
GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA  12420

  K  D  Y  P  N  L  L  N  A  I  N  H  L  I  L  S  K  T  S  N
AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT  12480

  C  N  D  F  I  L  L  I  A  G  D  G  A  L  R  N  K  L  L  D
TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT  12540

  L  V  C  Q  L  N  L  V  D  K  V  F  F  L  G  Q  R  S  D  I
TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT  12600
```

Figure 7/10

```
 K  E  L  M  C  A  A  D  L  F  V  L  S  S  E  W  E  G  F  G
AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT   12660

 L  V  V  A  E  A  M  A  C  E  R  P  V  V  A  T  D  S  G  G
CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA   12720

 V  K  E  V  V  G  P  H  N  D  V  I  P  V  S  N  H  I  L  L
GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG   12780

 A  E  K  I  A  E  T  L  K  I  D  D  N  A  R  K  I  I  G  M
GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG   12840

 K  N  R  E  Y  I  V  S  N  F  S  I  K  T  I  V  S  E  W  E
AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG   12900

                                                  End of orf11
 R  L  Y  F  K  Y  S  K  R  N  N  I  I  D  *
CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGATTGAAAATATAAGTTTGTA   12960

CTCTGGATGCAATAGTTTCTCTATGCTGTTTTTTTACTGGCTCCGTATTTTTACTTATAG   13020

CTGGATTTTGTTATATATCAGTATTAATCTGTCTCAACTTCATCTAGACTACATTCAAGC   13080

                                                  Start of gnd
                                                  M  S  K  Q  Q  I
CGCGCATGCGTCGCGCGGTGACTACACCTGACAGGAGTATGTAATGTCCAAGCAACAGAT   13140

 G  V  V  G  M  A  V  M  G  R  N  L  A  L  N  I  E  S  R  G
CGGCGTCGTCGGTATGGCAGTGATGGGGCGCAACCTGGCGCTCAACATCGAAAGCCGCGG   13200

 Y  T  V  S  I  F  N  R  S  R  E  K  T  E  E  V  V  A  E  N
TTATACCGTCTCCATCTTCAACCGCTCCCGCGAGAAAACTGAAGAAGTTGTTGCCGAGAA   13260

 P  D  K  K  L  V  P  Y  Y  T  V  K  E  F  V  E  S  L  E  T
CCCGGATAAGAAACTGGTTCCTTATTACACGGTGAAAGAGTTCGTCGAGTCTCTTGAAAC   13320

 P  R  R  I  L  L  M  V  K  A  G  A  G  T  D  A  A  I  D  S
CCCACGTCGTATCCTGTTAATGGTAAAAGCAGGGGCGGGAACTGATGCTGCTATCGATTC   13380

 L  K  P  Y  L  D  K  G  D  I  I  I  D  G  G  N  T  F  F  Q
CCTGAAGCCGTATCTGGATAAAGGCGACATCATTATTGATGGTGGCAACACCTTCTTCCA   13440

 D  T  I  R  R  N  R  E  L  S  A  E  G  F  N  F  I  G  T  G
GGACACTATCCGTCGTAACCGTGAACTGTCCGCGGAAGGCTTTAACTTCATCGGTACCGG   13500

 V  S  G  G  E  E  G  A  L  K  G  P  S  I  M  P  G  G  Q  K
CGTGTCCGGCGGTGAAGAGGGCGCCCTGAAAGGCCCATCTATCATGCCAGGTGGCCAGAA   13560

 E  A  Y  E  L  V  A  P  I  L  T  K  I  A  A  V  A  E  D  G
AGAAGCGTATGAGCTGGTTGCGCCTATCCTGACCAAGATTGCTGCGGTTGCTGAAGATGG   13620

 E  P  C  I  T  Y  I  G  A  D  G  A  G  H  Y  V  K  M  V  H
CGAACCATGTATAACTTACATCGGTGCTGACGGTGCGGGTCACTACGTGAAGATGGTGCA   13680

 N  G  I  E  Y  G  D  M  Q  L  I  A  E  A  Y  S  L  L  K  G
CAACGGTATCGAATATGGCGATATGCAGCTGATTGCTGAAGCCTATTCTCTGCTTAAAGG   13740

 G  L  N  L  S  N  E  E  L  A  T  T  F  T  E  W  N  E  G  E
CGGCCTTAATCTGTCTAACGAAGAGCTGGCAACCACTTTTACCGAGTGGAATGAAGGCGA   13800

 L  S  S  Y  L  I  D  I  T  K  D  I  F  T  K  K  D  E  E  G
GCTAAGTAGCTACCTGATTGACATCACCAAAGACATCTTCACCAAAAAAGATGAAGAGGG   13860
```

Figure 7/11

```
  K  Y  L  V  D  V  I  L  D  E  A  A  N  K  G  T  G  K  W  T
TAAATACCTGGTTGATGTGATCCTGGACGAAGCTGCGAACAAAGGCACCGGTAAATGGAC   13920

  S  Q  S  S  L  D  L  G  E  P  L  S  L  I  T  E  S  V  F  A
CAGCCAGAGCTCTCTGGATCTGGGTGAACCGCTGTCGCTGATCACCGAATCCGTATTCGC   13980

  R  Y  I  S  S  L  K  D  Q  R  I  A  A  S  K  V  L  S  G  P
TCGCTACATCTCTTCTCTGAAAGACCAGCGCATTGCGGCATCTAAAGTGCTGTCTGGTCC   14040

  Q  A  K  L  A  G  D  K  A  E  F  V  E  K  V  R  R  A  L  Y
GCAGGCTAAACTGGCTGGTGATAAAGCAGAGTTCGTTGAGAAAGTCCGTCGCGCGCTGTA   14100

  L  G  K  I  V  S  Y  A  Q  G  F  S  Q  L  R  A  A  S  D  E
CCTGGGTAAAATCGTCTCTTATGCCCAAGGCTTCTCTCAACTGCGTGCCGCGTCTGACGA   14160

  Y  N  W  D  L  N  Y  G  E  I  A  K  I  F  R  A  G  C  I  I
ATACAACTGGGATCTGAACTACGGCGAAATCGCGAAGATCTTCCGCGCGGGCTGCATCAT   14220

  R  A  Q  F  L  Q  K  I  T  D  A  Y  A  E  N  K  G  I  A  N
TCGTGCGCAGTTCCTGCAGAAAATTACTGACGCGTATGCTGAAAACAAAGGCATTGCTAA   14280

  L  L  A  P  Y  F  K  N  I  A  D  E  Y  Q  Q  A  L  R  D
CCTGTTGCTGGCTCCGTACTTCAAAAATATCGCTGATGAATATCAGCAAGCGCTGCGTGA   14340

  V  V  A  Y  A  V  Q  N  G  I  P  V  P  T  F  S  A  A  V  A
TGTAGTGGCTTATGCTGTGCAGAACGGTATTCCGGTACCGACCTTCTCTGCAGCGGTAGC   14400

  Y  Y  D  S  Y  R  S  A  V  L  P  A  N  L  I  Q  A  Q  R  D
CTACTACGACAGCTACCGTTCTGCGGTACTGCCGGCTAATCTGATTCAGGCACAGCGTGA   14460

  Y  F  G  A  H  T  Y  K  R  T  D  K  E  G  V  F  H  T  G
TTACTTCGGTGCGCACACGTATAAACGCACTGATAAAGAAGGTGTGTTCCACACCG   14516
```

Figure 7/12

```
GTAACCAAGGGCGGTACGTGCATAAATTTTAATGCTTATCAAAACTATTAGCATTAAAAA      60

                    Start of orf1
                  M  N  K  E  T  V  S  I  I  M  P  V  Y  N
TATATAAGAAATTCTCAAATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAAT     120

 G  A  K  T  I  I  S  S  V  E  S  I  I  H  Q  S  Y  Q  D  F
GGGGCCAAAACTATAATCTCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTT     180

 V  L  Y  I  I  D  D  C  S  T  D  D  T  F  S  L  I  N  S  R
GTTTTGTATATCATTGACGATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGA     240

 Y  K  N  N  Q  K  I  R  I  L  R  N  K  T  N  L  G  V  A  E
TACAAAAACAATCAGAAAATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAA     300

 S  R  N  Y  G  I  E  M  A  T  G  K  Y  I  S  F  C  D  A  D
AGTCGAAATTATGGAATAGAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGAT     360

 D  L  W  H  E  K  K  L  E  R  Q  I  E  V  L  N  N  E  C  V
GATTTGTGGCACGAGAAAAAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTA     420

 D  V  V  C  S  N  Y  Y  Y  V  I  D  N  N  R  N  I  V  G  E  V
GATGTGGTATGTTCTAATTATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTT     480

 N  A  P  H  V  I  N  Y  R  K  M  L  M  K  N  Y  I  G  N  L
AATGCTCCTCATGTGATAAATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTG     540

 T  G  I  Y  N  A  N  K  L  G  K  F  Y  Q  K  K  I  G  H  E
ACAGGAATCTATAATGCCAACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAG     600

 D  Y  L  M  W  L  E  I  I  N  K  T  N  G  A  I  C  I  Q  D
GATTATTTGATGTGGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGAT     660

 N  L  A  Y  Y  M  R  S  N  N  S  L  S  G  N  K  I  K  A  A
AATCTGGCGTATTACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCA     720

 K  W  T  W  S  I  Y  R  E  H  L  H  L  S  F  P  K  T  L  Y
AAATGGACATGGAGTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATAT     780

 Y  F  L  L  Y  A  S  N  G  V  M  K  K  I  T  H  S  L  L  R
TATTTTTTATTATATGCTTCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGG     840

                    Start of orf2, End of orf1
 R  K  E  T  K  K  *
                       V  K  S  A  A  K  L  I  F  L  F  L  F  T
AGAAAGGAGACTAAAAAGTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACAC     900

 L  Y  S  L  Q  L  Y  G  V  I  I  D  D  R  I  T  N  F  D  T
TTTATAGTCTCCAGTTGTATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAA     960

 K  V  L  T  S  I  I  I  I  F  Q  I  F  F  V  L  L  F  Y  L
AGGTATTAACTAGTATTATAATTATATTTCAGATTTTTTTTTGTTTTATTATTTTATCTAA    1020

 T  I  I  N  E  R  K  Q  Q  K  K  F  I  V  N  W  E  L  K  L
CGATTATAAATGAAAGAAAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAA    1080

 I  L  V  F  L  F  V  T  I  E  I  A  A  V  V  L  F  L  K  E
TACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAG    1140

 G  I  P  I  F  D  D  D  P  G  G  A  K  L  R  I  A  E  G  N
GTATTCCTATATTTGATGATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATG    1200
```

Figure 8/1

```
G  L  Y  I  R  Y  I  K  Y  F  G  N  I  V  V  F  A  L  I  I
GACTTTACATTAGATATATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTC    1260

L  Y  D  E  H  K  F  K  Q  R  T  I  I  F  V  Y  F  T  T  I
TTTATGATGAGCATAAATTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTG    1320

A  L  F  G  Y  R  S  E  L  V  L  L  I  L  Q  Y  I  L  I  T
CTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCA    1380

N  I  L  S  K  D  N  R  N  P  K  I  K  R  I  I  G  Y  F  L
ATATCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTAT    1440

L  V  G  V  V  C  S  L  F  Y  L  S  L  G  Q  D  G  E  Q  N
TGGTAGGGGTTGTATGCTCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATG    1500

D  S  Y  N  N  M  L  R  I  I  N  R  L  T  I  E  Q  V  E  G
ACTCATATAATAATATGTTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTG    1560

V  P  Y  V  V  S  E  S  I  K  N  D  F  F  P  T  P  E  L  E
TTCCATATGTTGTTTCTGAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAA    1620

K  E  L  K  A  I  I  N  R  I  Q  G  I  K  H  Q  D  L  F  Y
AGGAATTAAAAGCAATAATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATG    1680

G  E  R  L  H  K  Q  V  F  G  D  M  G  A  N  F  L  S  V  T
GAGAACGGTTACATAAACAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTA    1740

T  Y  G  A  E  L  L  V  F  F  G  F  L  C  V  F  I  I  P  L
CGTATGGAGCAGAACTGTTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAG    1800

G  I  Y  I  P  F  Y  L  L  K  R  M  K  K  T  H  S  S  I  N
GGATATATATACCTTTTTATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATT    1860

C  A  F  Y  S  Y  I  I  M  I  L  L  Q  Y  L  V  A  G  N  A
GCGCATTCTATTCATATATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCAT    1920

S  A  F  F  F  G  P  F  L  S  V  L  I  M  C  T  P  L  I  L
CGGCCTTCTTTTTTGGTCCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTAT    1980

                                Start of orf3
                              M  K  I  S  V  I  T  V  T  Y
L  H  D  T  L  K  R  L  S  R  N  E  N  I  S  Y  N  C  D  L
TGCATGATACGTTAAAGAGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTAT    2040

  End of orf2
  N  N  A  E  G  L  E  K  T  L  S  S  L  S  I  L  K  I  K  P
  *
AATAATGCTGAAGGGTTAGAAAAAACTTTAAGTAGTTTATCAATTTTAAAAATAAAACCT    2100

  F  E  I  I  I  V  D  G  G  S  T  D  G  T  N  R  V  I  S  R
TTTGAGATTATTATAGTTGATGGCGGCTCTACAGATGGAACGAATCGTGTCATTAGTAGA    2160

  F  T  S  M  N  I  T  H  V  Y  E  K  D  E  G  I  Y  D  A  M
TTTACTAGTATGAATATTACACATGTTTATGAAAAAGATGAAGGGATATATGATGCGATG    2220

  N  K  G  R  M  L  A  K  G  D  L  I  H  Y  L  N  A  G  D  S
AATAAGGGCCGAATGTTGGCCAAAGGCGACTTAATACATTATTTAAACGCCGGCGATAGC    2280

  V  I  G  D  I  Y  K  N  I  K  E  P  C  L  I  K  V  G  L  F
GTAATTGGAGATATATATAAAAATATCAAAGAGCCATGTTTGATTAAAGTTGGCCTTTTC    2340

  E  N  D  K  L  L  G  F  S  S  I  T  H  S  N  T  G  Y  C  H
GAAAATGATAAACTTCTGGGATTTTCTTCTATAACCCATTCAAATACAGGGTATTGTCAT    2400
```

Figure 8/2

```
 Q  G  V  I  F  P  K  N  H  S  E  Y  D  L  R  Y  K  I  C  A
CAAGGGGTGATTTTCCCAAAGAATCATTCAGAATATGATCTAAGGTATAAAATATGTGCT    2460

 D  Y  K  L  I  Q  E  V  F  P  E  G  L  R  S  L  S  L  I  T
GATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACT    2520

 S  G  Y  V  K  Y  D  M  G  G  V  S  S  K  K  R  I  L  R  D
TCGGGTTATGTAAAATATGATATGGGGGGAGTATCTTCAAAAAAAAGAATTTTAAGAGAT    2580

 K  E  L  A  K  I  M  F  E  K  N  K  K  N  L  I  K  F  I  P
AAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAAACCTTATTAAGTTTATTCCA    2640

 I  S  I  I  K  I  L  F  P  E  R  L  R  K  F  M  Q
ATTTCAATAATCAAAATTTTATTCCCTGAACGTTTAAGAAGAGTATTGCGGAAAATGCAA    2700
```

Start of orf4 End of orf3

```
 Y  I  C  L  T  L  F  F  M  K  N  S  S  P  Y  D  N  E  *
                                          M  I  M  N  K  I
TATATTTGTCTAACTTTATTCTTCATGAAGAATAGTTCACCATATGATAATGAATAAAAT    2760

 K  K  I  L  K  F  C  T  L  K  K  Y  D  T  S  S  A  L  G  R
CAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTGCTTTAGGTAG    2820

 E  Q  E  R  Y  R  I  I  S  L  S  V  I  S  S  L  I  S  K  I
AGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTAGTAAAAT    2880

 L  S  L  L  S  L  I  L  T  V  S  L  T  L  P  Y  L  G  Q  E
ACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATTTAGGACAAGA    2940

 R  F  G  V  W  M  T  I  T  S  L  G  A  A  L  T  F  L  D  L
GAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTGACATTTTTGGACTT    3000

 G  I  G  N  A  L  T  N  R  I  A  H  S  F  A  C  G  K  N  L
AGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTTGCGTGTGGCAAAAATTT    3060

 K  M  S  R  Q  I  S  G  G  L  T  L  L  A  G  L  S  F  V  I
AAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCTGGATTATCGTTTGTCAT    3120

 T  A  I  C  Y  I  T  S  G  M  I  D  W  Q  L  V  I  K  G  I
AACTGCAATATGCTATATTACTTCTGGCATGATTGATTGGCAACTAGTAATAAAAGGTAT    3180

 N  E  N  V  Y  A  E  L  Q  H  S  I  K  V  F  V  I  I  F  G
AAACGAGAATGTGTATGCAGAGTTACAACACTCAATTAAAGTCTTTGTAATCATATTTGG    3240

 L  G  I  Y  S  N  G  V  Q  K  V  Y  M  G  I  Q  K  A  Y  I
ACTTGGAATTTATTCAAATGGTGTGCAAAAAGTTTATATGGGAATACAAAAAGCCTATAT    3300

 S  N  I  V  N  A  I  F  I  L  L  S  I  I  T  L  V  I  S  S
AAGTAATATTGTTAATGCCATATTTATATTGTTATCTATTATTACTCTAGTAATATCGTC    3360

 K  L  H  A  G  L  P  V  L  I  V  S  T  L  G  I  Q  Y  I  S
GAAACTACATGCGGGACTACCAGTTTTAATTGTCAGCACTCTTGGTATTCAATACATATC    3420

 G  I  Y  L  T  I  N  L  I  I  K  R  L  I  K  F  T  K  V  N
GGGAATCTATTTAACAATTAATCTTATTATAAAGCGATTAATAAAGTTTACAAAAGTTAA    3480

 I  H  A  K  R  E  A  P  Y  L  I  L  N  G  F  F  F  F  I  L
CATACATGCTAAAAGAGAAGCTCCATATTTGATATTAAACGGTTTTTTCTTTTTATTTT    3540

 Q  L  G  T  L  A  T  W  S  G  D  N  F  I  I  S  I  T  L  G
ACAGTTAGGCACTCTGGCAACATGGAGTGGTGATAACTTTATAATATCTATAACATTGGG    3600
```


Figure 8/3

```
  V  T  Y  V  A  V  F  S  I  T  Q  R  L  F  Q  I  S  T  V  P
TGTTACTTATGTTGCTGTTTTTAGCATTACACAGAGATTATTTCAAATATCTACGGTCCC   3660

  L  T  I  Y  N  I  P  L  W  A  A  Y  A  D  A  H  A  R  N  D
TCTTACGATTTATAACATCCCGTTATGGGCTGCTTATGCAGATGCTCATGCACGCAATGA   3720

  T  Q  F  I  K  K  T  L  R  T  S  L  K  I  V  G  I  S  S  F
TACTCAATTTATAAAAAAGACGCTCAGAACATCATTGAAAATAGTGGGTATTTCATCATT   3780

  L  L  A  F  I  L  V  V  F  G  S  E  V  V  N  I  W  T  E  G
CTTATTGGCCTTCATATTAGTAGTGTTCGGTAGTGAAGTCGTTAATATTTGGACAGAAGG   3840

  K  I  Q  V  P  R  T  F  I  I  A  Y  A  L  W  S  V  I  D  A
AAAGATTCAGGTACCTCGAACATTCATAATAGCTTATGCTTTATGGTCTGTTATTGATGC   3900

  F  S  N  T  F  A  S  F  L  N  G  L  N  I  V  K  Q  Q  M  L
TTTTTCGAATACATTTGCAAGCTTTTTAAATGGTTTGAACATAGTTAAACAACAAATGCT   3960

  A  V  V  T  L  I  L  I  A  I  P  A  K  Y  I  I  V  S  H  F
TGCTGTTGTAACATTGATATTGATCGCAATTCCAGCAAAATACATCATAGTTAGCCATTT   4020

  G  L  T  V  M  L  Y  C  F  I  F  I  Y  I  V  N  Y  F  I  W
TGGGTTAACTGTTATGTTGTACTGCTTCATTTTTATATATATTGTAAATTACTTTATATG   4080

                                  Start of orf5, End of orf4
                                        M  K  M
  Y  K  C  S  F  K  K  H  I  D  R  Q  L  N  I  R  G  *
GTATAAATGTAGTTTTAAAAAACATATCGATAGACAGTTAAATATAAGAGGATGAAAATG   4140

  K  Y  I  P  V  Y  Q  P  S  L  T  G  K  E  K  E  Y  V  N  E
AAATATATACCAGTTPACCAACCGTCATTGACAGGAAAAGAAAAAGAATATGTAAATGAA   4200

  C  L  D  S  T  W  I  S  S  K  G  N  Y  I  Q  K  F  E  N  K
TGTCTGGACTCAACGTGGATTTCATCAAAAGGAAACTATATTCAGAAGTTTGAAAATAAA   4260

  F  A  E  Q  N  H  V  Q  Y  A  T  T  V  S  N  G  T  V  A  L
TTTGCGGAACAAAACCATGTGCAATATGCAACTACTGTAAGTAATGGAACGGTTGCTCTT   4320

  H  L  A  L  L  A  L  G  I  S  E  G  D  E  V  I  V  P  T  L
CATTTAGCTTTGTTAGCGTTAGGTATATCGGAAGGAGATGAAGTTATTGTTCCAACACTG   4380

  T  Y  I  A  S  V  N  A  I  K  Y  T  G  A  T  P  I  F  V  D
ACATATATAGCATCAGTTAATGCTATAAAATACACAGGAGCCACCCCCATTTTCGTTGAT   4440

  S  D  N  E  T  W  Q  M  S  V  S  D  I  E  Q  K  I  T  N  K
TCAGATAATGAAACTTGGCAAATGTCTGTTAGTGACATAGAACAAAAAATCACTAATAAA   4500

  T  K  A  I  M  C  V  H  L  Y  G  H  P  C  D  M  E  Q  I  V
ACTAAAGCTATTATGTGTGTCCATTTATACGGACATCCATGTGATATGGAACAAATTGTA   4560

  E  L  A  K  S  R  N  L  F  V  I  E  D  C  A  E  A  F  G  S
GAACTGGCCAAAAGTAGAAATTTGTTTGTAATTGAAGATTGCGCTGAAGCCTTTGGTTCT   4620

  K  Y  K  G  K  Y  V  G  T  F  G  D  I  S  T  F  S  F  F  G
AAATATAAAGGTAAATATGTGGGAACATTTGGAGATATTTCTACTTTTAGCTTTTTTGGA   4680

  N  K  T  I  T  T  G  E  G  G  M  V  V  T  N  D  K  T  L  Y
AATAAAACTATTACTACAGGTGAAGGTGGAATGGTTGTCACGAATGACAAAACACTTTAT   4740

  D  R  C  L  H  F  K  G  Q  G  L  A  V  H  R  Q  Y  W  H  D
GACCGTTGTTTACATTTTAAAGGCCAAGGATTAGCTGTACATAGGCAATATTGGCATGAC   4800

  V  I  G  Y  N  Y  R  M  T  N  I  C  A  A  I  G  L  A  Q  L
GTTATAGGCTACAATTATAGGATGACAAATATCTGCGCTGCTATAGGATTAGCCCAGTTA   4860
```


Figure 8/4

```
E  Q  A  D  D  F  I  S  R  K  R  E  I  A  D  I  Y  K  K  N
GAACAAGCTGATGATTTTATATCACGAAAACGTGAAATTGCTGATATTTATAAAAAAAT  4920

I  N  S  L  V  Q  V  H  K  E  S  K  D  V  F  H  T  Y  W  M
ATCAACAGTCTTGTACAAGTCCACAAGGAAAGTAAAGATGTTTTTCACACTTATTGGATG  4980

V  S  I  L  T  R  T  A  E  E  R  E  E  L  R  N  H  L  A  D
GTCTCAATTCTAACTAGGACCGCAGAGGAAAGAGAGGAATTAAGGAATCACCTTGCAGAT  5040

K  L  I  E  T  R  P  V  F  Y  P  V  H  T  M  P  M  Y  S  E
AAACTCATCGAAACAAGGCCAGTTTTTTACCCTGTCCACACGATGCCAATGTACTCGGAA  5100

K  Y  Q  K  H  P  I  A  E  D  L  G  W  R  G  I  N  L  P  S
AAATATCAAAAGCACCCTATAGCTGAGGATCTTGGTTGGCGTGGAATTAATTTACCTAGT  5160

F  P  S  L  S  N  E  Q  V  I  Y  I  C  E  S  I  N  E  F  Y
TTCCCCAGCCTATCGAATGAGCAAGTTATTTATATTTGTGAATCTATTAACGAATTTTAT  5220

        End of orf5                    Start of orf6
S  D  K  *                          M  K  I  A  L  N  S  D
AGTGATAAATAGCCTAAAATATTGTAAAGGTCATTCATGAAAATTGCGTTGAATTCAGAT  5280

G  F  Y  E  W  G  G  G  I  D  F  I  K  Y  I  L  S  I  L  E
GGATTTTACGAGTGGGGCGGTGGAATTGATTTTATTAAATATATTCTGTCAATATTAGAA  5340

T  K  P  E  I  C  I  D  I  L  L  P  R  N  D  I  H  S  L  I
ACGAAACCAGAAATATGTATCGATATTCTTTTACCGAGAAATGATATACATTCTCTTATA  5400

R  E  K  A  F  P  F  K  S  I  L  K  A  I  L  K  R  E  R  P
AGAGAAAAAGCATTTCCTTTTAAAAGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCT  5460

R  W  I  S  L  N  R  F  N  E  Q  Y  Y  R  D  A  F  T  Q  N
CGATGGATTTCATTAAATAGATTTAATGAGCAATACTATAGAGATGCCTTTACACAAAAT  5520

N  I  E  T  N  L  T  F  I  K  S  K  S  S  A  F  Y  S  Y  F
AATATAGAGACGAATCTTACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTT  5580

D  S  S  D  C  D  V  I  L  P  C  M  R  V  P  S  G  N  L  N
GATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAAT  5640

K  K  A  W  I  G  Y  I  Y  D  F  Q  H  C  Y  Y  P  S  F  F
AAAAAAGCATGGATTGGTTATATTTATGACTTTCAACACTGTTACTATCCTTCATTTTTT  5700

S  K  R  E  I  D  Q  R  N  V  F  F  K  L  M  L  N  C  A  N
AGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTAAATTGATGCTCAATTGCGCTAAC  5760

N  I  I  V  N  A  H  S  V  I  T  D  A  N  K  Y  V  G  N  Y
AATATTATTGTTAATGCACATTCAGTTATTACCGATGCAAATAAATATGTTGGGAATTAT  5820

S  A  K  L  H  S  L  P  F  S  P  C  P  Q  L  K  W  F  A  D
TCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGAT  5880

Y  S  G  N  I  A  K  Y  N  I  D  K  D  Y  F  I  I  C  N  Q
TACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTATAATTTGCAATCAA  5940

F  W  K  H  K  D  H  A  T  A  F  R  A  F  K  I  Y  T  E  Y
TTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATAT  6000

N  P  D  V  Y  L  V  C  T  G  A  T  Q  D  Y  R  F  P  G  Y
AATCCTGATGTTTATTTAGTATGCACGGGAGCTACTCAAGATTATCGATTCCCTGGATAT  6060

F  N  E  L  M  V  L  A  K  K  L  G  I  E  S  K  I  K  I  L
TTTAATGAATTGATGGTTTTGGCAAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTA  6120
```


Figure 8/5

```
 G  H  I  P  K  L  E  Q  I  E  L  I  K  N  C  I  A  V  I  Q
GGGCATATACCTAAACTTGAACAAATTGAATTAATCAAAAATTGCATTGCTGTAATACAA   6180

 P  T  L  F  E  G  G  P  G  G  G  V  T  F  D  A  I  A  L  G
CCAACCTTATTTGAAGGCGGGCCTGGAGGGGGGGTAACATTTGACGCTATTGCATTAGGG   6240

 K  K  V  I  L  S  D  I  D  V  N  K  E  V  N  C  G  D  V  Y
AAAAAAGTTATACTATCTGACATAGATGTCAATAAAGAAGTTAATTGCGGTGATGTATAT   6300

 F  F  Q  A  K  N  H  Y  S  L  N  D  A  M  V  K  A  D  E  S
TTCTTTCAGGCAAAAAACCATTATTCATTAAATGACGCGATGGTAAAAGCTGATGAATCT   6360

 K  I  F  Y  E  P  T  T  L  I  E  L  G  L  K  R  R  N  A  C
AAAATTTTTTATGAACCTACAACTCTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGT   6420

                                                   End of orf6
 A  D  F  L  L  D  V  V  K  Q  E  I  E  S  R  S  *
GCAGATTTTCTTTTAGATGTTGTGAAACAAGAAATTGAATCCCGATCT TAATATATTCAA   6480

                Start of orf7
                M  T  K  V  A  L  I  T  G  V  T  G  Q  D  G  S  Y
GAGGTATATAATGACTAAAGTCGCTCTTATTACAGGTGTAACTGGACAAGATGGATCTTA   6540

 L  A  E  F  L  L  D  K  G  Y  E  V  H  G  I  K  R  R  A  S
TCTAGCTGAGTTTTTGCTTGATAAAGGGTATGAAGTTCATGGTATCAAACGCCGAGCCTC   6600

 S  F  N  T  E  R  I  D  H  I  Y  Q  D  P  H  G  S  N  P  N
ATCTTTTAATACAGAACGCATAGACCATATTTATCAAGATCCACATGGTTCTAACCCAAA   6660

 F  H  L  H  Y  G  D  L  T  D  S  S  N  L  T  R  I  L  K  E
TTTTCACTTGCACTATGGAGATCTGACTGATTCATCTAACCTCACTAGAATTCTAAAGGA   6720

 V  Q  P  D  E  V  Y  N  L  A  A  M  S  H  V  A  V  S  F  E
GGTACAGCCAGATGAAGTATATAATTTAGCTGCTATGAGTCACGTAGCAGTTTCTTTTGA   6780

 S  P  E  Y  T  A  D  V  D  A  I  G  T  L  R  L  L  E  A  I
GTCTCCAGAATATACAGCCGATGTCGATGCAATTGGTACATTACGTTTACTGGAAGCAAT   6840

 R  F  L  G  L  E  N  K  T  R  F  Y  Q  A  S  T  S  E  L  Y
TCGCTTTTTAGGATTGGAAAACAAAACGCGTTTCTATCAAGCTTCAACCTCAGAATTATA   6900

 G  L  V  Q  E  I  P  Q  K  E  S  T  P  F  Y  P  R  S  P  Y
TGGACTTGTTCAGGAAATCCCTCAAAAAGAATCCACCCCTTTTTATCCTCGTTCCCCTTA   6960

 A  V  A  K  L  Y  A  Y  W  I  T  V  N  Y  R  E  S  Y  G  I
TGCAGTTGCAAAACTTTACGCATATTGGATCACGGTAAATTATCGAGAGTCATATGGTAT   7020

 Y  A  C  N  G  I  L  F  N  H  E  S  P  R  R  G  E  T  F  V
TTATGCATGTAATGGTATATTGTTCAATCATGAATCTCCACGCCGTGGAGAAACGTTTGT   7080

 T  R  K  I  T  R  G  L  A  N  I  A  Q  G  L  E  S  C  L  Y
AACAAGGAAAATTACTCGAGGACTTGCAAATATTGCACAAGGCTTGGAATCATGTTTGTA   7140

 L  G  N  M  D  S  L  R  D  W  G  H  A  K  D  Y  V  R  M  Q
TTTAGGGAATATGGATTCGTTACGAGATTGGGGACATGCAAAAGATTATGTTAGAATGCA   7200

 W  L  M  L  Q  Q  E  Q  P  E  D  F  V  I  A  T  G  V  Q  Y
ATGGTTGATGTTACAACAGGAGCAACCCGAAGATTTTGTGATTGCAACAGGAGTCCAATA   7260

 S  V  R  Q  F  V  E  M  A  A  A  Q  L  G  I  K  M  S  F  V
CTCAGTCCGTCAGTTTGTCGAAATGGCAGCAGCACAACTTGGTATTAAGATGAGCTTTGT   7320
```


Figure 8/6

```
  G  K  G  I  E  E  K  G  I  V  D  S  V  E  G  Q  D  A  P  G
TGGTAAAGGAATCGAAGAAAAAGGCATTGTAGATTCGGTTGAAGGACAGGATGCTCCAGG      7380

  V  K  P  G  D  V  I  V  A  V  D  P  R  Y  F  R  P  A  E  V
TGTGAAACCAGGTGATGTCATTGTTGCTGTTGATCCTCGTTATTTCCGACCAGCTGAAGT      7440

  D  T  L  L  G  D  P  S  K  A  N  L  K  L  G  W  R  P  E  I
TGATACTTTGCTTGGAGATCCGAGCAAAGCTAATCTCAAACTTGGTTGGAGACCAGAAAT      7500

  T  L  A  E  M  I  S  E  M  V  A  K  D  L  E  A  A  K  K  H
TACTCTTGCTGAAATGATTTCTGAAATGGTTGCCAAAGATCTTGAAGCCGCTAAAAAACA      7560
```

Start of orf8, End of orf7

```
                                                M  M  M  N  K
  S  L  L  K  S  H  G  F  S  V  S  L  A  L  E  *
TTCTCTTTTAAAATCGCATGGTTTTTCTGTAAGCTTAGCTCTGGAATGATGATGAATAAG      7620

  Q  R  I  F  I  A  G  H  Q  G  M  V  G  S  A  I  T  R  R  L
CAACGTATTTTTATTGCTGGTCACCAAGGAATGGTTGGATCAGCTATTACCCGACGCCTC      7680

  K  Q  R  D  D  V  E  L  V  L  R  T  R  D  E  L  N  L  L  D
AAACAACGTGATGATGTTGAGTTGGTTTTACGTACTCGGGATGAATTGAACTTGTTGGAT      7740

  S  S  A  V  L  D  F  F  S  S  Q  K  I  D  Q  V  Y  L  A  A
AGTAGCGCTGTTTTGGATTTTTTTTCTTCACAGAAAATCGACCAGGTTTATTTGGCAGCA      7800

  A  K  V  G  G  I  L  A  N  S  S  Y  P  A  D  F  I  Y  E  N
GCAAAAGTCGGAGGTATTTAGCTAACAGTTCTTTATCCTGCCGATTTTATATATGAGAAT      7860

  I  M  I  E  A  N  V  I  H  A  A  H  K  N  N  V  N  K  L  L
ATAATGATAGAGGCGAATGTCATTCATGCTGCCCACAAAAATAATGTAAATAAACTGCTT      7920

  F  L  G  S  S  C  I  Y  P  K  L  A  H  Q  P  I  M  E  D  E
TTCCTCGGTTCGTCGTGTATTTATCCTAAGTTAGCACACCAACCGATTATGGAAGACGAA      7980

  L  L  Q  G  K  L  E  P  T  N  E  P  Y  A  I  A  K  I  A  G
TTATTACAAGGGAAACTTGAGCCAACAAATGAACCTTATGCTATCGCAAAAATTGCAGGT      8040

  I  K  L  C  E  S  Y  N  R  Q  F  G  R  D  Y  R  S  V  M  P
ATTAAATTATGTGAATCTTATAACCGTCAGTTTGGGCGTGATTACCGTTCAGTAATGCCA      8100

  T  N  L  Y  G  P  N  D  N  F  H  P  S  N  S  H  V  I  P  A
ACCAATCTTTATGGTCCAAATGACAATTTTCATCCAAGTAATTCTCATGTGATTCCGGCG      8160

  L  L  R  R  F  H  D  A  V  E  N  N  S  P  N  V  V  V  W  G
CTTTTGCGCCGCTTTCATGATGCTGTGGAAAACAATTCTCCGAATGTTGTTGTTTGGGGA      8220

  S  G  T  P  K  R  E  F  L  H  V  D  D  M  A  S  A  S  I  Y
AGTGGTACTCCAAAGCGTGAATTCTTACATGTAGATGATATGGCTTCTGCAAGCATTTAT      8280

  V  M  E  M  P  Y  D  I  W  Q  K  N  T  K  V  M  L  S  H  I
GTCATGGAGATGCCATACGATATATGGCAAAAAAATACTAAAGTAATGTTGTCTCATATC      8340

  N  I  G  T  G  I  D  C  T  I  C  E  L  A  E  T  I  A  K  V
AATATTGGAACAGGTATTGACTGCACGATTTGTGAGCTTGCGGAAACAATAGCAAAAGTT      8400

  V  G  Y  K  G  H  I  T  F  D  T  T  K  P  D  G  A  P  R  K
GTAGGTTATAAAGGGCATATTACGTTCGATACAACAAAGCCCGATGGAGCCCCTCGAAAA      8460

  L  L  D  V  T  L  L  H  Q  L  G  W  N  H  K  I  T  L  H  K
CTACTTGATGTAACGCTTCTTCATCAACTAGGTTGGAATCATAAAATTACCCTTCACAAG      8520
```

Figure 8/7

End of orf8

G L E N T Y N W F L E N Q L Q Y R G *
GGTCTTGAAAATACATACAACTGGTTTCTTGAAAACCAACTTCAATATCGGGGG TAATAA 8580

Start of orf9

M F L H S Q D F A T I V R S T P L I S I
TGTTTTTACATTCCCAAGACTTTGCCACAATTGTAAGGTCTACTCCTCTTATTTCTATAG 8640

D L I V E N E F G E I L L G K R I N R P
ATTTGATTGTGGAAAACGAGTTTGGCGAAATTTTGCTAGGAAAACGAATCAACCGCCCGG 8700

A Q G Y W F V P G G R V L K D E K L Q T
CACAGGGCTATTGGTTCGTTCCTGGTGGTAGGGTGTTGAAAGATGAAAAATTGCAGACAG 8760

A F E R L T E I E L G I R L P L S V G K
CCTTTGAACGATTGACAGAAATTGAACTAGGAATTCGTTTGCCTCTCTCTGTGGGTAAGT 8820

F Y G I W Q H F Y E D N S M G G D F S T
TTTATGGTATCTGGCAGCACTTCTACGAAGACAATAGTATGGGGGGAGACTTTTCAACGC 8880

H Y I V I A F L L K L Q P N I L K L P K
ATTATATAGTTATAGCATTCCTTCTTAAATTACAACCAAACATTTTGAAATTACCGAAGT 8940

S Q H N A Y C W L S R A K L I N D D D V
CACAACATAATGCTTATTGCTGGCTATCGCGAGCAAAGCTGATAAATGATGACGATGTGC 9000

H Y N C R A Y F N N K T N D A I G L D N
ATTATAATTGTCGCGCATATTTTAACAATAAAACAAATGATGCGATTGGCTTAGATAATA 9060

Start of orf10 **End of orf9**

M S D A P I I A V V M A G G T G S
K D I I C L M R Q *
AGGATATAATATGTCTGATGCGCCAATAATTGCTGTAGTTATGGCCGGTGGTACAGGCAG 9120

R L W P L S R E L Y P K Q F L Q L S G D
TCGTCTTTGGCCACTTTCTCGTGAACTATATCCAAAGCAGTTTTTACAACTCTCTGGTGA 9180

N T L L Q T T L L R L S G L S C Q K P L
TAACACCTTGTTACAAACGACTTTGCTACGACTTTCAGGCCTATCATGTCAAAAACCATT 9240

V I T N E Q H R F V V A E Q L R E I N K
AGTGATAACAAATGAACAGCATCGCTTTGTTGTGGCTGAACAGTTAAGGGAAATAAATAA 9300

L N G N I I L E P C G R N T A P A I A I
ATTAAATGGTAATATTATTCTAGAACCATGCGGGCGAAATACTGCACCAGCAATAGCGAT 9360

S A F H A L K R N P Q E D P L L L V L A
ATCTGCGTTTCATGCGTTAAAACGTAATCCTCAGGAAGATCCATTGCTTCTAGTTCTTGC 9420

A D H V I A K E S V F C D A I K N A T P
GGCAGACCACGTTATAGCTAAAGAAAGTGTTTTCTGTGATGCTATTAAAAATGCAACTCC 9480

I A N Q G K I V T F G I I P E Y A E T G
CATCGCTAATCAAGGTAAAATTGTAACGTTTGGAATTATACCAGAATATGCTGAAACTGG 9540

Y G Y I E R G E L S V P L Q G H E N T G
TTATGGGTATATTGAGAGAGGTGAACTATCTGTACCGCTTCAAGGGCATGAAAATACTGG 9600

F Y Y V N K F V E K P N R E T A E L Y M
TTTTTATTATGTAAATAAGTTTGTCGAAAAGCCTAATCGTGAAACCGCAGAATTGTATAT 9660

T S G N H Y W N S G I F M F K A S V Y L
GACTTCTGGTAATCACTATTGGAATAGTGGAATATTCATGTTTAAGGCATCTGTTTATCT 9720

Figure 8/8

```
  E  E  L  R  K  F  R  P  D  I  Y  N  V  C  E  Q  V  A  S  S
TGAGGAATTGAGAAAATTTAGACCTGACATTTACAATGTTTGTGAACAGGTTGCCTCATC    9780

  S  Y  I  D  L  D  F  I  R  L  S  K  E  Q  F  Q  D  C  P  A
CTCATACATTGATCTAGATTTTATTCGATTATCAAAAGAACAATTTCAAGATTGTCCTGC    9840

  E  S  I  D  F  A  V  M  E  K  T  E  K  C  V  V  C  P  V  D
TGAATCTATTGATTTTGCTGTAATGGAAAAAACAGAAAAATGTGTTGTATGCCCTGTTGA    9900

  I  G  W  S  D  V  G  S  W  Q  S  L  W  D  I  S  L  K  S  K
TATTGGTTGGAGTGACGTTGGATCTTGGCAATCGTTATGGGACATTAGTCTAAAATCGAA    9960

  T  G  D  V  C  K  G  D  I  L  T  Y  D  T  K  N  N  Y  I  Y
AACAGGAGATGTATGTAAAGGTGATATATTAACCTATGATACTAAGAATAATTATATCTA    10020

  S  E  S  A  L  V  A  A  I  G  I  E  D  M  V  I  V  Q  T  K
CTCTGAGTCAGCGTTGGTAGCCGCCATTGGAATTGAAGATATGGTTATCGTGCAAACTAA    10080

  D  A  V  L  V  S  K  K  S  D  V  Q  H  V  K  K  I  V  E  M
AGATGCCGTTCTTGTGTCTAAAAAGAGTGATGTACAGCATGTAAAAAAAATAGTCGAAAT    10140

  L  K  L  Q  Q  R  T  E  Y  I  S  H  R  E  V  F  R  P  W  G
GCTTAAATTGCAGCAACGTACAGAGTATATTAGTCATCGTGAAGTTTTCCGACCATGGGG    10200

  K  F  D  S  I  D  Q  G  E  R  Y  K  V  K  K  I  I  V  K  P
AAAATTTGATTCGATTGACCAAGGTGAGCGATACAAAGTCAAGAAAATTATTGTGAAACC    10260

  G  E  G  L  S  L  R  M  H  H  H  R  S  E  H  W  I  V  L  S
TGGTGAGGGGCTTTCTTTAAGGATGCATCACCATCGTTCTGAACATTGGATCGTGCTTTC    10320

  G  T  A  K  V  T  L  G  D  K  T  K  L  V  T  A  N  E  S  I
TGGTACAGCAAAAGTAACCCTTGGCGATAAAACTAAACTAGTCACCGCAAATGAATCGAT    10380

  Y  I  P  L  G  A  A  Y  S  L  E  N  P  G  I  I  P  L  N  L
ATACATTCCCCTTGGCGCAGCGTATAGTCTTGAGAATCCGGGCATAATCCCTCTTAATCT    10440

  I  E  V  S  S  G  D  Y  L  G  E  D  D  I  I  R  Q  K  E  R
TATTGAAGTCAGTTCAGGGGATTATTTGGGAGAGGATGATATTATAAGACAGAAAGAACG    10500

                End of orf10  Start of orf11
  Y  K  H  E  D  *     M  K  S  L  T  C  F  K  A  Y  D  I  R
TTACAAACATGAAGATTAACATATGAAATCTTTAACCTGCTTTAAAGCCTATGATATTCG    10560

  G  K  L  G  E  E  L  N  E  D  I  A  W  R  I  G  R  A  Y  G
CGGGAAATTAGGCGAAGAACTGAATGAAGATATTGCCTGGCGCATTGGGCGTGCCTATGG    10620

  E  F  L  K  P  K  T  I  V  L  G  G  D  V  R  L  T  S  E  A
CGAATTTCTCAAACCGAAAACCATTGTTTAGGCGGTGATGTCCGCCTCACCAGCGAAGC    10680

  L  K  L  A  L  A  K  G  L  Q  D  A  G  V  D  V  L  D  I  G
GTTAAAACTGGCGCTTGCGAAAGGTTTACAGGATGCGGGCGTCGATGTGCTGGATATCGG    10740

  M  S  G  T  E  E  I  Y  F  A  T  F  H  L  G  V  D  G  G  I
TATGTCCGGCACCGAAGAGATCTATTTCGCCACGTTCCATCTCGGAGTGGATGGCGGCAT    10800

  E  V  T  A  S  H  N  P  M  D  Y  N  G  M  K  L  V  R  E  G
CGAAGTTACCGCCAGCCATAACCCGATGGATTACAACGGCATGAAGCTGGTGCGCGAAGG    10860

  A  R  P  I  S  G  D  T  G  L  R  D  V  Q  R  L  A  E  A  N
GGCTCGCCCGATCAGCGGTGATACCGGACTGCGCGATGTCCAGCGTCTGGCAGAAGCCAA    10920

  D  F  P  P  V  D  E  T  K  R  G  R  Y  Q  Q  I  N  L  R  D
TGACTTCCCTCCTGTCGATGAAACCAAACGTGGTCGCTATCAGCAAATCAATCTGCGTGA    10980
```


Figure 8/9

```
  A  Y  V  D  H  L  F  G  Y  I  N  V  K  N  L  T  P  L  K  L
CGCTTACGTTGATCACCTGTTCGGTTATATCAACGTCAAAAACCTCACGCCGCTCAAGCT   11040

  V  I  N  S  G  N  G  A  A  G  P  V  V  D  A  I  E  A  R  F
GGTGATCAACTCCGGGAACGGCGCAGCGGGTCCGGTGGTGGACGCCATTGAAGCCCGATT   11100

  K  A  L  G  A  P  V  E  L  I  K  V  H  N  T  P  D  G  N  F
TAAAGCCCTCGGCGCACCGGTGGAATTAATCAAAGTACACAACACGCCGGACGGCAATTT   11160

  P  N  G  I  P  N  P  L  L  P  E  C  R  D  D  T  R  N  A  V
CCCCAACGGTATTCCTAACCCGCTGCTGCCGGAATGCCGCGACGACACCCGTAATGCGGT   11220

  I  K  H  G  A  D  M  G  I  A  F  D  G  D  F  D  R  C  F  L
CATCAAACACGGCGCGGATATGGGCATTGCCTTTGATGGCGATTTTGACCGCTGTTTCCT   11280

  F  D  E  K  G  Q  F  I  E  G  Y  Y  I  V  G  L  L  A  E  A
GTTTGACGAAAAAGGGCAGTTTATCGAGGGCTACTACATTGTCGGCCTGCTGGCAGAAGC   11340

  F  L  E  K  N  P  G  A  K  I  I  H  D  P  R  L  S  W  N  T
GTTCCTCGAAAAAAATCCCGGCGCGAAGATCATCCACGATCCACGTCTCTCCTGGAACAC   11400

  V  D  V  V  T  A  A  G  G  T  P  V  M  S  K  T  G  H  A  F
CGTTGATGTGGTGACTGCCGCAGGCGGCACCCCGGTAATGTCGAAAACCGGACACGCCTT   11460

  I  K  E  R  M  R  K  E  D  A  I  Y  G  G  E  M  S  A  H  H
TATTAAAGAACGTATGCGCAAGGAAGACGCCATCTACGGTGGCGAAATGAGCGCTCACCA   11520

  Y  F  R  D  F  A  Y  C  D  S  G  M  I  P  W  L  L  V  A  E
TTACTTCCGTGATTTCGCTTACTGCGACAGCGGCATGATCCCGTGGCTGCTGGTCGCCGA   11580

  L  V  C  L  K  G  K  T  L  G  E  M  V  R  D  R  M  A  A  F
ACTGGTGTGCCTGAAAGGAAAAACGCTGGGCGAAATGGTGCGCGACCGGATGGCGGCGTT   11640

  P  A  S  G  E  I  N  S  K  L  A  Q  P  V  E  A  I  N  R  V
TCCGGCAAGCGGTGAGATCAACAGCAAACTGGCGCAACCCGTTGAGGCAATTAATCGCGT   11700

  E  Q  H  F  S  R  E  A  L  A  V  D  R  T  D  G  I  S  M  T
GGAACAGCATTTTAGCCGCGAGGCGCTGGCGGTGGATCGCACCGATGGCATCAGCATGAC   11760

  F  A  D  W  R  F  N  L  R  S  S  N  T  E  P  V  V  R  L  N
CTTTGCCGACTGGCGCTTTAACCTGCGCTCCTCCAACACCGAACCGGTGGTGCGGTTGAA   11820

  V  E  S  R  G  D  V  K  L  M  E  K  K  T  K  A  L  L  K  L
TGTGGAATCACGCGGTGATGTAAAGCTAATGGAAAAGAAAACTAAAGCTCTTCTTAAATT   11880

                End of orf11
  L  S  E  *
GCTAAGTGAGTGATTATTTACATTAATCATTAAGCGTATTTAAGATTATATTAAAGTAAT   11940

GTTATTGCGGTATATGATGAATATGTGGGCTTTTTTATGTATAACGACTATACCGCAACT   12000

          Start of R-repeat
TTATCTAGGAAAAGATTAATAGAAATAAAGTTTTGTACTGACCAATTTGCATTTCACGTC   12060

ACGATTGAGACGTTCCTTTGCTTAAGACATTTTTTCATCGCTTATGTAATAACAAATGTG   12120

CCTTATATAAAAAGGAGAACAAAATGGAACTTAAAATAATTGAGACAATAGATTTTTATT   12180

ATCCCTGTTTACGATATTATAGCCAAAGTTGTATCCTGCATCAGTCCTGCAATATTTCAC   12240

GAGTGCTTTGTTAACTGAATACATGTCTGCCATTTTCCAGATGATAACGACGTCATCGCA   12300

ATTGATGGTAAAACACTTCGGCACACTTATGACAAGAGTCGTCGCAGAGGAGTGGTTCAT   12360
```

Figure 8/10

```
GTCATTAGTGCGTTTCAGCAATGCACAGTCTGGTCCTCGGATAGATCAAGACGGATGAGA   12420

AACCTAATGCGTTCACAGTTATTCATGAACTTTCTAAAATGATGGGTATTAAAGGAAAAA   12480

TAATCATAACTGATGCGATGGCTTGCCAGAAAGATATTGCAGAGAAGATATAAAAACAGA   12540

GATGTGATTATTTATTCGCTGTAAAAGGAAATAAGAGTCGGCTTAATAGAGTCTTTGAGG   12600

AGATATTTACGCTGAAAGAATTAAATAATCCAAAACATGACAGTTACGCAATTAGTGAAA   12660

AGAGGCACGGCAGAGACGATGTCCGTCTTCATATTGTTTGAGATGCTCCTGATGAGCTTA   12720

TTGATTTCACGTTTGAATGGAAAGGGCTGCAGAATTTATGAATGGCAGTCCACTTTCTCT   12780

CAATAATAGCAGAGCAAAAGAAAGAATCCGAAATGACGATCAAATATTATATTAGATCTG   12840

CTGCTTTAACCGCAGAGAAGTTCGCCACAGTAAATCGAAATCACTGGCGCATGGAGAATA   12900

AGTTGCACAGTAGCCTGATGTGGTAATGAATGAAATCGACTATAATATAAGAAGGCGAGT   12960

TGCATTCGAATGATTTTCTAGAATGCGGCACATCGCTATTAATATCTGACAATGATAATG   13020

TATTCAAGGCAGGATTATCATGTAAGATGCGAAAAGCAGTCATGGACAGAAACTTCCTAG   13080

                                          End of the H-repeat
CGTCAGGCATTGCAGCGTGCGGGCTTTCATAATCTTGCAT TGGTTTTGATAAGATATTTC   13140

          Start of orf12
            M  N  L  Y  G  I  F  G  A  G  S  Y  G  R  E
TTTGGAGATGGGAAAATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAA   13200

 T  I  P  I  L  N  Q  Q  I  K  Q  E  C  G  S  D  Y  A  L  V
ACAATACCCATTCTAAATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTT   13260

 F  V  D  D  V  L  A  G  K  K  V  N  G  F  E  V  L  S  T  N
TTTGTGGATGATGTTTTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAAC   13320

 C  F  L  K  A  P  Y  L  K  K  Y  F  N  V  A  I  A  N  D  K
TGCTTTCTAAAAGCCCCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAG   13380

 I  R  Q  R  V  S  E  S  I  L  L  H  G  V  E  P  I  T  I  K
ATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAA   13440

 H  P  N  S  V  V  Y  D  H  T  M  I  G  S  G  A  I  I  S  P
CATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCC   13500

 F  V  T  I  S  T  N  T  H  I  G  R  F  F  H  A  N  I  Y  S
TTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCA   13560

 Y  V  A  H  D  C  Q  I  G  D  Y  V  T  F  A  P  G  A  K  C
TACGTTGCACATGATTGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGT   13620

 N  G  Y  V  V  I  E  D  N  A  Y  I  G  S  G  A  V  I  K  Q
AATGGATATGTTGTTATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAG   13680

 G  V  P  N  R  P  L  I  I  G  A  G  A  I  I  G  M  G  A  V
GGTGTTCCTAATCGCCCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTT   13740

 V  T  K  S  V  P  A  G  I  T  V  C  G  N  P  A  R  E  M  K
GTCACTAAAAGTGTTCCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAA   13800

                End of orf12
 R  S  P  T  S  I  *
AGATCGCCAACATCTATT TAATGGGAATGCGAAAACACGTTCCAAATGGGACTAATGTTT   13860
```

Figure 8/11

AAAATATATATAATTTCGCTAATTTACTAAATTATGGCTTCTTTTTAAGCTATCCTTTAC 13920

TTAGTTATTACTGATACAGCATGAAATTTATAATACTCTGATACATTTTTATACGTTATT 13980

CAAGCCGCATATCTAGCGGTAACCCCTGACAGGAGTAAACAATG 14024

Figure 8/12

```
GTTGACAAATACCGACCGTATAATGAATCAAACGTTCTGGATTGGTATTTATCCAGGCTT                60

GACTACAGAGCATTTAGATTATGTCGTAAGTAAGTTTGAAGAATTTTTTGGTTTAAATTT               120

                Start of abe
                M  L  D  V  N  K  K  I  L  M  T  G  A  T
CTAATTTTTAGGATAGGATGCTTGATGTGAATAAGAAAATCCTAATGACTGGCGCTACTA               180

S  F  V  G  T  H  L  L  H  S  L  I  K  E  G  Y  S  I  I  A
GCTTTGTAGGTACCCATCTACTACATAGTCTCATAAAGGAAGGTTATAGTATTATTGCAT               240

L  K  R  P  I  T  E  P  T  I  I  N  T  L  I  E  W  L  N  I
TAAAGCGTCCTATAACCGAGCCAACGATTATCAATACCTTGATTGAATGGTTGAATATAC               300

Q  D  I  E  K  I  C  Q  S  S  M  N  I  H  A  I  V  H  I  A
AAGATATAGAAAAAATATGTCAATCATCTATGAATATTCATGCGATTGTCCATATTGCAA               360

T  D  Y  G  R  N  R  T  P  I  S  E  Q  Y  K  C  N  V  L  L
CAGACTATGGTCGAAACAGAACCCTATATCTGAACAATATAAATGTAATGTCCTATTAC               420

P  T  R  L  L  E  L  M  P  A  L  K  T  K  F  F  I  S  T  D
CAACAAGACTGCTTGAGTTAATGCCAGCGCTTAAAACGAAATTCTTTATTTCTACTGACT              480

S  F  F  G  K  Y  E  K  H  Y  G  Y  M  R  S  Y  M  A  S  K
CTTTTTTTGGGAAATATGAGAAGCACTATGGATATATGCGTTCTTACATGGCATCTAAAA              540

R  H  F  V  E  L  S  K  I  Y  V  E  E  H  P  D  V  C  F  I
GACATTTTGTAGAACTATCAAAAATATACGTAGAGGAACATCCAGACGTTTGTTTTATAA              600

N  L  R  L  E  H  V  Y  G  E  R  D  K  A  G  K  I  I  P  Y
ATTTACGTTTAGAACATGTTTACGGTGAGAGGGATAAAGCAGGTAAAATAATCCCGTATG              660

V  I  K  K  M  K  N  N  E  D  I  D  C  T  I  A  R  Q  K  R
TTATCAAAAAAATGAAAAACAATGAAGATATTGATTGTACGATCGCCAGGCAGAAAAGAG              720

D  F  I  Y  I  D  D  V  V  S  A  Y  L  K  I  L  K  E  G  F
ATTTTATTTATATAGACGATGTTGTTTCGGCCTATTTGAAAATTTTAAAGGAGGGTTTTA              780

N  A  G  H  Y  D  V  E  V  G  T  G  K  S  I  E  L  K  E  V
ACGCTGGACACTATGATGTCGAGGTGGGGACTGGAAAATCGATAGAGCTAAAAGAAGTGT              840

F  E  I  I  K  K  E  T  H  S  S  S  K  I  N  Y  G  A  V  A
TTGAGATAATAAAAAAAGAAACGCATAGTAGTAGTAAGATAAATTATGGTGCAGTTGCGA              900

M  R  D  D  E  I  M  E  S  H  A  N  T  S  F  L  T  R  L  G
TGCGTGATGATGAGATTATGGAGTCACATGCAAATACCTCTTTCTTGACTCGATTAGGTT              960

                                    End of abe  Start of wzx
                                                        M
W  S  A  E  F  S  I  E  K  G  V  K  K  M  L  S  M  K  E  *
GGAGTGCCGAGTTTTCTATTGAGAAGGGTGTGAAAAAAATGTTGAGTATGAAAGAG TAAT             1020

  N  R  I  I  R  M  L  G  V  D  K  A  I  R  Y  V  I  F  G  K
GAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGTAA             1080

  I  I  S  V  L  T  G  L  L  L  I  M  L  I  S  H  H  L  S  K
GATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCTAA             1140

  D  A  Q  G  Y  Y  Y  T  F  N  S  V  V  A  L  Q  I  I  F  E
AGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTTGA             1200
```

Figure 9/1

```
  L  G  L  S  T  V  I  I  Q  F  A  S  H  E  M  S  A  L  K  Y
ATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAATA                1260

  D  Y  S  E  R  D  I  I  G  E  S  K  N  K  Q  R  Y  L  S  L
TGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCGTT                1320

  F  R  L  A  I  K  W  Y  A  V  I  A  L  L  I  I  L  I  V  G
ATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTCGG                1380

  P  I  G  Y  V  F  F  T  Q  K  E  G  L  G  V  P  W  Q  G  A
TCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGGGC               1440

  W  L  L  L  T  I  V  T  A  F  N  I  F  L  V  S  V  L  S  V
ATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGT                1500

  A  E  G  S  G  L  I  T  D  V  N  K  M  R  M  Y  Q  S  L  L
CGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTT               1560

  A  G  I  L  A  V  S  L  L  I  S  G  F  G  L  Y  A  T  S  A
AGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGC                1620

  I  A  I  S  G  T  I  I  F  S  I  F  S  Y  K  Y  F  K  K  I
AATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAAT                1680

  F  L  Q  S  L  K  H  K  N  K  Y  T  E  G  G  I  S  W  V  N
TTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAA                1740

  E  I  F  P  M  Q  W  R  I  A  L  S  W  M  S  G  Y  F  I  Y
TGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTA                1800

  F  V  M  T  P  I  A  F  K  Y  F  G  A  I  Y  A  G  Q  L  G
TTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTAGG                1860

  M  S  L  T  L  C  N  M  V  M  A  T  G  L  A  W  I  S  T  K
GATGTCTTTAACATTATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACTAA                1920

  Y  P  K  W  G  V  M  V  S  N  K  Q  L  A  E  L  S  K  S  F
ATATCCAAAATGGGGAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCGTT                1980

  K  S  A  V  M  Q  S  S  F  F  V  L  T  G  L  T  G  V  Y  I
CAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATACAT                2040

  S  L  W  L  L  K  L  S  G  S  N  I  G  E  R  F  L  G  L  Q
TTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTGCA                2100

  D  F  F  F  L  S  L  A  I  I  G  N  H  I  V  A  C  F  A  T
GGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCAAC                2160

  Y  I  R  A  H  K  T  E  K  M  T  L  A  S  C  I  M  A  L  L
CTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGCTCTCTT                2220

  T  I  T  T  M  L  F  V  A  Y  L  E  Y  S  R  F  Y  M  L  M
GACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACATGTTAAT                2280

  Y  A  A  L  T  W  L  Y  F  V  P  Q  T  Y  I  I  P  K  R  F
                                                S  L  K  D
GTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATCTTTAAAAGATT                2340
```

Figure 9/2

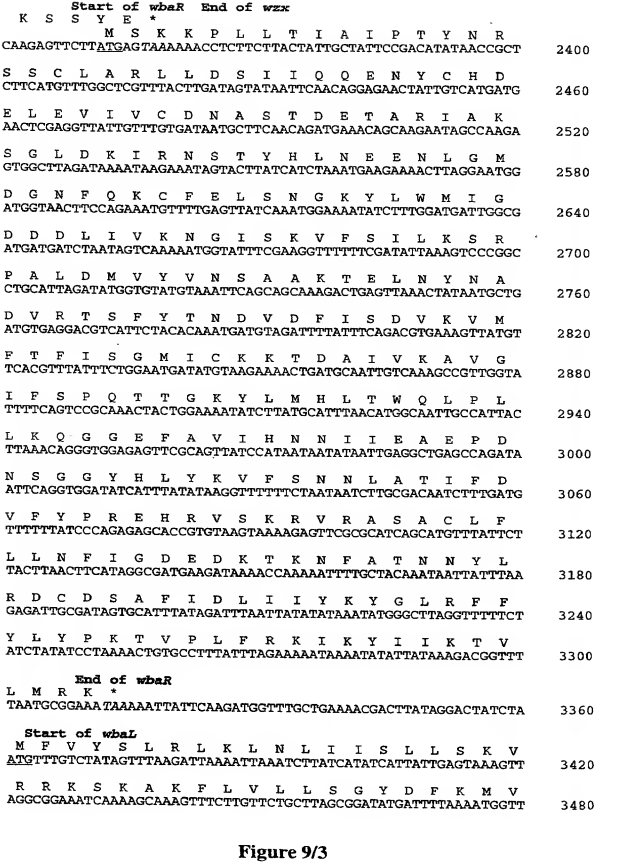
Start of wbaR End of wzx
K S S Y E *
M S K K P L L T I A I P T Y N R
CAAGAGTTCTTATGAGTAAAAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCT 2400
S S C L A R L L D S I I Q Q E N Y C H D
CTTCATGTTTGGCTCGTTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATG 2460
E L E V I V C D N A S T D E T A R I A K
AACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGA 2520
S G L D K I R N S T Y H L N E E N L G M
GTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGG 2580
D G N F Q K C F E L S N G K Y L W M I G
ATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCG 2640
D D D L I V K N G I S K V F S I L K S R
ATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGC 2700
P A L D M V Y V N S A A K T E L N Y N A
CTGCATTAGATATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATAATGCTG 2760
D V R T S F Y T N D V D F I S D V K V M
ATGTGAGGACGTCATTCTACACAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGT 2820
F T F I S G M I C K K T D A I V K A V G
TCACGTTTATTTCTGGAATGATATGTAAGAAAACTGATGCAATTGTCAAAGCCGTTGGTA 2880
I F S P Q T T G K Y L M H L T W Q L P L
TTTTCAGTCCGCAAACTACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTAC 2940
L K Q G G E F A V I H N N I I E A E P D
TTAAACAGGGTGGAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATA 3000
N S G G Y H L Y K V F S N N L A T I F D
ATTCAGGTGGATATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATG 3060
V F Y P R E H R V S K R V R A S A C L F
TTTTTTATCCCAGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCT 3120
L L N F I G D E D K T K N F A T N N Y L
TACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAA 3180
R D C D S A F I D L I I Y K Y G L R F F
GAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCT 3240
Y L Y P K T V P L F R K I K Y I I K T V
ATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTT 3300
End of wbaR
L M R K *
TAATGCGGAAATAAAAATTATTCAAGATGGTTTGCTGAAAACGACTTATAGGACTATCTA 3360
Start of wbaL
M F V Y S L R L K L N L I I S L L S K V
ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 3420
R R K S K A K F L V L L S G Y D F K M V
AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 3480

Figure 9/3

```
 G  K  N  F  K  L  N  V  K  P  Y  S  A  K  N  N  T  S  S  K
GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA    3540

 W  G  S  M  R  V  G  D  N  C  W  I  E  A  V  Y  N  Y  G  D
TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGAT    3600

 E  K  F  E  P  Y  L  Y  I  G  D  R  I  C  L  S  D  N  V  H
GAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCAT    3660

 I  S  C  V  S  C  L  I  L  E  N  D  I  L  I  G  S  K  V  Y
ATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTAT    3720

 I  G  D  H  S  H  G  S  Y  K  V  C  S  P  K  I  E  P  P  A
ATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCA    3780

 N  K  P  L  G  D  I  A  P  I  K  I  G  N  C  C  W  I  G  D
AATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGATTGGAGAT    3840

 N  A  V  I  L  A  G  S  E  I  C  D  G  C  V  I  A  A  N  S
AATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGATGGCTGTGTAATCGCAGCTAATTCA    3900

 V  V  K  D  L  K  V  D  K  P  C  L  I  G  G  V  P  A  K  V
GTCGTCAAGGATTTAAAAGTCGATAAGCCATGTTTAATTGGTGGGGTTCCTGCTAAAGTA    3960

                End of wbaL  Start of wbaQ
 I  K  V  F  *
                        M  N  V  F  I  S  I  C  I  P  S  Y  N  R  A
ATAAAGGTATTTTAAAATGAATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGC    4020

 E  F  L  E  P  L  L  D  S  I  Y  N  Q  D  Y  C  L  K  N  N
TGAGTTTTTAGAGCCACTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAA    4080

 D  F  E  V  I  V  C  E  D  K  S  P  Q  R  D  E  I  N  S  I
TGATTTTGAGGTCATTGTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTAT    4140

 I  E  N  Y  K  A  K  N  N  K  Q  N  L  Y  V  N  F  N  E  D
TATCGAAAACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGA    4200

 N  L  G  Y  D  K  N  L  K  K  C  I  S  L  T  T  G  K  Y  C
TAATTTAGGCTATGATAAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTG    4260

 M  I  M  G  N  D  D  L  L  A  D  G  A  L  S  K  I  V  K  V
CATGATCATGGGCAACGATGATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGT    4320

 L  K  A  N  P  E  I  V  L  A  T  R  A  Y  G  W  F  K  E  N
TTTGAAGGCTAATCCTGAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAA    4380

 P  N  E  L  C  D  T  V  R  H  L  T  D  D  T  L  F  Q  P  G
TCCGAATGAGTTATGTGATACTGTTCGTCATTTAACAGACGATACTTTATTTCAGCCGGG    4440

 A  D  A  I  K  F  F  F  R  R  V  G  V  I  S  G  F  I  V  N
GGCTGATGCCATTAAATTTTTCTTCCGTAGAGTTGGAGTTATTTCAGGCTTTATTGTCAA    4500

 A  E  K  A  K  K  L  S  S  D  L  F  D  G  R  L  Y  Y  Q  M
TGCTGAAAAAGCAAAAAAACTATCGAGTGATTTATTTGATGGGCGTTTATATTATCAAAT    4560

 Y  L  A  G  M  L  M  A  E  G  Q  G  Y  Y  F  S  D  V  M  T
GTACCTTGCTGGTATGCTAATGGCTGAAGGTCAGGGATACTATTTTAGCGACGTGATGAC    4620
```

Figure 9/4

```
  L  S  R  D  T  E  A  P  D  F  G  N  A  G  T  E  K  G  V  F
ATTGTCGAGGGATACAGAGGCTCCTGACTTTGGTAACGCTGGAACTGAAAAAGGAGTTTT  4680

  T  P  G  G  Y  K  P  E  G  R  I  H  M  V  E  G  L  L  L  I
CACCCCGGGGGGGTATAAACCAGAGGGCCGTATACATATGGTTGAAGGCTTGTTGCTAAT  4740

  A  K  Y  I  E  D  T  T  K  I  D  G  V  Y  A  G  I  R  K  D
TGCAAAATATATAGAAGATACAACAAAAATTGATGGCGTTTATGCTGGAATTAGAAAAGA  4800

  L  A  N  Y  F  Y  P  Y  I  R  D  Q  L  D  L  P  L  Y  T  Y
CTTAGCGAACTATTTTTATCCTTATATTCGAGATCAACTCGACTTGCCTCTTTATACTTA  4860

  I  K  M  I  N  K  F  R  K  M  G  F  S  N  E  K  L  F  Y  V
TATTAAAATGATAAATAAATTTCGGAAAATGGGATTTTCAAATGAAAAGCTTTTCTATGT  4920

  H  A  F  L  G  Y  V  L  K  R  R  G  Y  D  A  L  I  K  Y  I
GCATGCCTTTTTAGGGTATGTACTAAAACGGAGGGGCTATGATGCTTTAATTAAATACAT  4980

                                End of wbaQ
  R  S  K  K  G  G  T  P  R  L  G  I  *
TCGTAGCAAAAAAGGCGGTACTCCGCGTCTTGGTATTTAACCTCCACTTTCAAAAAATGT  5040

TATGAATATACTTCTTGCTGCGATATTAGGCGTTAACTTATTTTCTCCATATATTAGTTC  5100

                Start of wzy
                M  L  P  F  F  P  P  G  A  I  L  R  D  V  L  N  V
GTGGATGGTGGGTATGCTGCCATTTCCACCAGGAGCAATCCTAAGGGATGTACTCAATGT  5160

  F  F  V  A  L  V  L  V  R  F  V  I  D  R  K  K  T  Y  F  P
ATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCCC  5220

  L  V  F  T  I  F  S  W  S  A  V  I  L  W  V  I  A  L  T  I
GTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTAT  5280

  F  S  P  D  K  I  Q  A  I  M  G  G  R  S  Y  I  L  F  P  A
ATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGC  5340

  V  F  I  A  L  V  I  L  K  V  S  Y  P  Q  S  L  N  I  E  K
AGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAA  5400

  I  V  C  Y  I  I  F  L  M  F  M  V  A  T  I  S  I  I  D  V
AATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGT  5460

  L  M  N  G  E  F  I  K  L  L  G  Y  D  E  H  Y  A  G  E  Q
ACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACA  5520

  L  N  L  I  N  S  Y  D  G  M  V  R  A  T  G  G  F  S  D  A
ATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTAGTGATGC  5580

  L  N  F  G  Y  M  L  T  L  G  V  L  L  C  M  E  C  F  S  Q
TCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCA  5640

  G  Y  K  R  L  L  M  L  I  I  S  F  V  L  F  I  A  I  C  M
AGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCAT  5700

  S  L  T  R  G  A  I  L  V  A  A  L  I  Y  A  L  Y  I  I  S
GAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTC  5760

  N  R  K  M  L  F  C  G  I  T  L  F  V  I  I  I  P  V  L  A
AAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGC  5820
```

Figure 9/5

```
  I  S  T  N  I  F  D  N  Y  T  E  I  L  I  G  R  F  T  D  S
AATTTCTACTAATATTTTTGACAACTATACAGAAATTTTGATCGGCAGGTTTACAGATTC   5880

  S  Q  A  S  R  G  S  T  Q  G  R  I  D  M  A  I  N  S  L  N
GTCTCAGGCATCGCGTGGATCTACACAGGGGCGGATAGATATGGCAATTAATTCATTAAA   5940

  F  L  S  E  H  P  S  G  I  G  L  G  T  Q  G  S  G  N  M  L
CTTCCTGTCAGAACATCCATCAGGTATAGGTCTGGGTACTCAAGGTTCAGGAAACATGCT   6000

  S  V  K  D  N  R  L  N  T  D  N  Y  F  F  W  I  A  L  E  T
TTCGGTAAAAGATAATAGGTTAAATACGGATAATTATTTTTTCTGGATCGCCCTTGAGAC   6060

  G  I  I  G  L  I  I  N  I  I  Y  L  A  S  Q  F  Y  S  S  T
TGGTATTATTGGCTTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAAC   6120

  L  L  N  R  I  Y  G  S  H  C  S  N  M  H  Y  R  L  Y  F  L
TTTACTAAATAGAATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCT   6180

  F  G  S  I  Y  F  I  S  A  A  L  S  S  A  P  S  S  S  T  F
CTTTGGAAGTATATATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTT   6240

  S  I  Y  Y  W  T  V  L  A  L  I  P  F  L  K  L  T  N  R  R
TTCTATATATTATTGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAACAAATAGACG   6300

  End of wzy Start of wbaW
  C  T  R  *  M  N  N  K  K  V  L  M  D  I  S  W  S  N  K  G
GTGCACGCGATAATGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGG   6360

  G  I  G  R  F  T  D  E  I  S  K  L  L  C  D  I  S  K  E  E
GGGATTGGACGTTTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAA   6420

  L  Y  R  K  C  A  S  P  L  A  P  L  G  L  A  V  N  I  F  L
CTATATAGAAAATGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTG   6480

  R  K  T  D  V  V  F  L  P  G  Y  I  P  P  L  F  C  S  K
CGAAAGAAAACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAA   6540

  K  F  I  I  T  I  H  D  L  N  H  L  D  L  N  D  N  S  S  L
AAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTT   6600

  F  K  R  L  F  Y  N  F  I  I  K  R  G  C  R  K  A  Y  K  I
TTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATA   6660

  F  T  V  S  N  F  S  K  E  R  I  V  A  W  S  G  V  N  P  N
TTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAAT   6720

  K  I  V  T  V  Y  N  G  V  S  S  L  F  N  A  D  V  K  P  L
AAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTG   6780

  N  L  G  Y  K  Y  L  L  C  V  G  N  R  K  T  H  K  N  E  K
AATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAG   6840

  C  V  I  S  A  F  A  K  A  D  I  D  P  S  I  K  L  V  F  T
TGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACT   6900

  G  N  P  C  N  D  L  E  K  L  I  I  Q  H  G  L  S  E  R  V
GGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTA   6960
```

Figure 9/6

```
K F F G F V S E K D L P S L Y K G S L G
AAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGA 7020

L V F P S L Y E G F G L P V V E G M A C
TTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGT 7080

G I P V L T S L T S S L P E V A G D A A
GGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCG 7140

I L V D P L S E D A I T K G I S R L I N
ATTCTTGTCGACCCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAAT 7200

D S E L R K H L I Q K G L L R A K R F N
GATTCTGAACTTCGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAAT 7260

                                    Start of wbaZ
W Q N V V S E I E M V L T E A C D G N K
                                        M E I N
TGGCAAAACGTGGTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTGATGGAAATAAA 7320

End of wbaW
*
E I K I S L V H E W L L S Y A G S E Q V
TGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGCTCCGAACAGGT 7380

S S A I L H V F P E A K L Y S V V D F L
ATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTGGTTGATTTTCT 7440

T D E Q R R H F L G K Y A T T T F I Q N
AACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACATTTATTCAAAA 7500

L P K A K K F Y Q K Y L P L M P L A I E
TTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACTGGCTATTGA 7560

Q L D L S D A N I I I S S A H S V A K G
ACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGTTGCAAAAGG 7620

V I S G P D Q L H I S Y V H S P I R Y A
TGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCTATTCGATATGC 7680

W D L Q H Q Y L N E S N L N K G I K G W
GTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTAAAGGTTG 7740

L A K W L L H K I R I W D S R T A N G V
GTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACCGCAAATGGGGT 7800

D H F I A N S Q Y I A R R I K K V Y R R
TGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACG 7860

E A S V I Y P P V D V D N F E V K N E K
TGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAA 7920

Q D Y Y F T A S R M V P Y K R I D L I V
GCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGT 7980

E A F S K M P E K K L V V I G D G P E M
CGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGAT 8040

K K I K S K A T D N I K L L G Y Q S F P
GAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTATCAATCTTTTCC 8100
```

Figure 9/7

```
  V  L  K  E  Y  M  Q  S  A  R  A  F  V  F  A  A  E  E  D  F
TGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTT   8160

  G  I  I  P  V  E  A  Q  A  C  G  T  P  V  I  A  F  G  K  G
TGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGG   8220

  G  A  L  E  T  V  R  P  L  G  V  E  E  P  T  G  I  F  F  K
TGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGCATTTTCTTCAA   8280

  E  Q  N  I  A  S  L  H  E  A  V  S  E  F  E  K  N  A  S  F
GGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAAAATGCATCATT   8340

  F  T  S  Q  A  C  R  K  N  A  E  K  F  S  R  S  R  F  E  Q
TTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCAAGATTTGAACA   8400

  E  F  K  N  F  V  N  E  K  W  N  L  F  K  T  E  Q  I  I  K
AGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGAACAGATTATTAA   8460

  End of wbaZ Start of manC
                M  S  K  L  I  P  V  I  M  A  G  G  I
  R  *
ACGTTAATTATGGTTTATTGAATGTCTAAATTAATACCAGTAATAATGGCCGGTGGGATT   8520

  G  S  R  L  W  P  L  S  R  E  E  H  P  K  Q  F  L  S  V  D
GGTAGCCGTTTGTGGCCACTTTCACGTGAAGAGCATCCGAAACAGTTTTTAAGCGTAGAT   8580

  G  E  L  S  M  L  Q  N  T  I  K  R  L  T  P  L  L  A  G  E
GGTGAATTATCTATGCTGCAAAACACCATTAAAAGATTGACTCCTCTTTTGGCTGGAGAA   8640

  P  L  V  I  C  N  D  S  H  R  F  L  V  A  E  Q  L  R  A  I
CCTTTAGTCATTTGTAATGATAGTCACCGCTTCCTTGTCGCTGAACAACTTCGAGCTATA   8700

  N  K  L  A  N  N  I  I  L  E  P  V  G  R  N  T  A  P  A  I
AATAAACTAGCAAATAACATCATATTAGAGCCAGTGGGGCGTAATACAGCCCCAGCTATA   8760

  A  L  A  A  F  C  S  L  Q  N  V  V  D  E  D  P  L  L  L  V
GCGCTGGCCGCTTTTTGTTCACTTCAGAATGTCGTCGATGAAGACCCGCTTTTGCTTGTC   8820

  L  A  A  D  H  V  I  R  D  E  K  V  F  L  K  A  I  N  H  A
CTTGCTGCGGATCATGTCATCCGCGATGAGAAAGTGTTTCTTAAAGCTATCAATCACGCT   8880

  E  F  F  A  T  Q  G  K  L  V  T  F  G  I  V  P  T  Q  A  E
GAATTTTTTGCAACACAAGGTAAGCTAGTAACGTTTGGTATTGTACCCACACAGGCCGAA   8940

  T  G  Y  G  Y  I  C  R  G  E  A  I  G  E  D  A  F  S  V  A
ACTGGCTACGGTTATATTTGTAGAGGTGAAGCAATCGGGGAAGATGCTTTTTCTGTAGCC   9000

  E  F  V  E  K  P  D  F  D  T  A  R  H  Y  V  E  S  E  K  Y
GAATTTGTAGAGAAGCCTGATTTCGATACAGCGCGTCATTATGTAGAATCAGAGAAATAT   9060

  Y  W  N  S  G  M  F  L  F  R  A  S  S  Y  L  Q  E  L  K  D
TATTGGAACAGCGGTATGTTCCTATTTCGTGCAAGTAGTTACTTACAAGAATTAAAGGAT   9120

  L  S  P  D  I  Y  Q  A  C  E  N  A  V  G  S  I  N  P  D  L
CTGTCCCCCGATATTTACCAAGCATGTGAAAATGCGGTAGGGAGTATTAATCCTGATCTT   9180

  D  F  I  R  I  D  K  E  A  F  A  M  C  P  S  D  S  I  D  Y
GATTTTATCCGTATTGATAAAGAAGCATTCGCAATGTGCCCTAGTGATTCTATCGATTAT   9240
```

Figure 9/8

```
 A  V  M  E  H  T  R  H  A  V  V  V  P  M  N  A  G  W  S  D
GCGGTAATGGAACATACTAGGCATGCAGTTGTCGTACCGATGAATGCCGGCTGGTCAGAT   9300

 V  G  S  W  S  S  L  W  D  I  S  K  K  D  P  Q  R  N  V  L
GTGGGGTCATGGTCTTCACTGTGGGATATTTCTAAGAAAGATCCACAACGTAATGTATTA   9360

 H  G  D  I  F  A  Y  N  S  K  D  N  Y  I  Y  S  E  K  S  F
CATGGCGATATTTTTGCATATAATAGTAAAGATAATTATATCTATTCTGAAAAATCGTTT   9420

 I  S  T  I  G  V  N  N  L  V  I  V  Q  T  A  D  A  L  L  V
ATTAGTACAATCGGAGTAAATAATTTAGTTATCGTGCAGACAGCAGATGCATTATTAGTA   9480

 S  D  K  D  S  V  Q  D  V  K  K  V  V  D  Y  L  K  A  N  N
TCTGATAAAGATTCAGTCCAGGATGTTAAAAAAGTTGTTGATTATTTAAAAGCTAATAAT   9540

 R  N  E  H  K  K  H  L  E  V  F  R  P  W  G  K  F  S  V  I
AGAAACGAACATAAAAAACATTTAGAGGTTTTCCGACCGTGGGGAAAATTTAGCGTAATT   9600

 H  S  G  D  N  Y  L  V  K  R  I  T  V  K  P  G  A  K  F  A
CATAGTGGCGATAATTATTTAGTTAAAAGAATAACTGTTAAACCAGGCGCGAAGTTTGCT   9660

 A  Q  M  H  L  H  R  A  E  H  W  I  V  V  S  G  T  A  C  I
GCTCAGATGCATCTCCATCGTGCTGAGCATTGGATAGTGGTATCTGGTACTGCTTGTATT   9720

 T  K  G  E  E  I  F  T  I  S  E  N  E  S  T  F  I  P  A  N
ACTAAGGGGGAAGAAATTTTTACAATTTCGGAGAATGAATCAACATTTATACCTGCTAAT   9780

 T  V  H  T  L  K  N  P  A  T  I  P  L  E  L  I  E  I  Q  S
ACAGTTCATACGTTAAAAAACCCCGCGACTATTCCATTAGAACTAATAGAAATTCAATCT   9840

 G  T  Y  L  A  E  D  D  I  I  R  L  E  K  H  S  G  Y  L  E
GGCACCTATCTTGCGGAGGATGATATTATTCGCCTGGAGAAACATTCTGGATATCTGGAG   9900

  End of manC Start of manB
  *
           M  K  N  I  Y  N  T  Y  D  V  I  N  K  S  G  I  N
TAATGAATTGATGAAAAATATATATAATACTTACGATGTTATCAACAAATCTGGAATTAA   9960

 F  G  T  S  G  A  R  G  L  V  T  D  F  T  P  E  V  C  A  R
TTTTGGAACCAGTGGTGCCCGCGGCCTTGTTACCGATTTTACACCCGAAGTTTGCGCACG  10020

 F  T  I  S  F  L  T  V  M  Q  Q  R  F  S  F  T  T  V  A  L
ATTTACCATTTCCTTTTTGACAGTAATGCAGCAAAGATTCTCATTTACAACGGTTGCGCT  10080

 A  I  D  N  R  P  S  S  Y  A  M  A  Q  A  C  A  A  A  L  Q
CGCAATTGATAATCGTCCAAGCAGTTACGCGATGGCTCAAGCTTGTGCCGCTGCTTTGCA  10140

 E  K  G  I  K  T  V  Y  Y  G  V  I  P  T  P  A  L  A  H  Q
AGAAAAAGGAATTAAAACCGTTTACTATGGCGTAATTCCAACACCTGCTTTAGCTCATCA  10200

 S  I  S  D  K  V  P  A  I  M  V  T  G  S  H  I  P  F  D  R
ATCAATTTCCGATAAAGTACCTGCAATCATGGTTACTGGCAGTCATATCCCTTTTGACCG  10260

 N  G  L  K  F  Y  R  P  D  G  E  I  T  K  D  D  E  N  A  I
TAATGGCCTGAAATTTTATAGACCAGATGGTGAAATTACTAAAGATGATGAGAATGCTAT  10320

 I  H  V  D  A  S  F  M  Q  P  K  L  E  Q  L  T  I  S  T  I
TATTCATGTTGATGCCTCATTTATGCAGCCTAAGCTTGAACAATTGACAATTTCCACAAT  10380
```

Figure 9/9

```
  A  A  R  N  Y  I  L  R  Y  T  S  L  F  P  M  P  F  L  K  N
CGCTGCTAGAAATTATATTCTACGATATACCTCATTATTTCCAATGCCATTCTTGAAAAA   10440

  K  R  I  G  I  Y  E  H  S  S  A  G  R  D  L  Y  K  T  L  F
TAAGCGCATTGGAATTTATGAGCATTCTAGTGCGGGTCGTGATCTCTATAAGACGTTATT   10500

  K  M  L  G  A  T  V  V  S  L  A  R  S  D  E  F  V  P  I  D
CAAAATGTTGGGTGCTACAGTTGTTAGTTTAGCAAGGAGCGACGAATTTGTTCCTATTGA   10560

  T  E  A  V  S  E  D  D  R  N  K  A  I  T  W  A  K  K  Y  Q
TACTGAAGCTGTAAGTGAAGATGATAGAAATAAAGCAATCACATGGGCAAAAAAATATCA   10620

  L  D  A  I  F  S  T  D  G  D  G  D  R  P  L  I  A  D  E  Y
GTTAGATGCTATATTTTCAACTGATGGTGATGGAGATCGCCCTCTGATAGCTGACGAATA   10680

  G  N  W  L  R  G  D  I  L  G  L  L  C  S  L  E  L  A  A  D
TGGAAATTGGTTAAGAGGAGATATTATTAGGCCTTCTGTGCTCTCTCGAATTAGCTGCTGA   10740

  A  V  A  I  P  V  S  C  N  S  T  I  S  S  G  N  F  F  K  H
TGCAGTCGCTATTCCTGTAAGCTGCAACAGTACAATCTCATCTGGTAACTTTTTTAAACA   10800

  V  E  R  T  K  I  G  S  P  Y  V  I  A  A  F  A  K  L  S  A
TGTGGAACGAACAAAGATTGGTTCACCCTATGTGATTGCAGCATTTGCTAAATTATCTGC   10860

  N  Y  N  C  I  A  G  F  E  A  N  G  G  F  L  L  G  S  D  V
AAACTATAATTGTATAGCTGGTTTTGAAGCGAATGGTGGCTTTCTGCTAGGTAGCGATGT   10920

  Y  I  N  Q  R  L  L  K  A  L  P  T  R  D  A  L  L  P  A  I
TTATATTAATCAGCGTTTACTTAAGGCATTACCAACACGTGATGCTTTATTACCTGCCAT   10980

  M  L  L  F  G  S  K  D  K  S  I  S  E  L  V  K  K  L  P  A
TATGCTTCTGTTTGGTAGCAAGGACAAAAGTATTAGTGAGCTTGTTAAAAAACTTCCTGC   11040

  R  Y  T  Y  S  N  R  L  Q  D  I  S  V  K  T  S  M  S  L  I
TCGCTATACCTATTCAAACAGATTACAGGATATAAGTGTTAAAACAAGTATGTCTTTAAT   11100

  N  L  G  L  T  D  Q  E  D  F  L  Q  Y  I  G  F  N  K  H  H
AAATCTTGGTCTGACAGATCAAGAGGATTTTTTGCAGTATATTGGTTTTAATAAACATCA   11160

  I  L  H  S  D  V  T  D  G  F  R  I  T  I  D  N  N  N  I  I
TATATTACATTCTGATGTTACTGATGGCTTTAGAATCACTATCGATAACAACAATATTAT   11220

  H  L  R  P  S  G  N  A  P  E  L  R  C  Y  A  E  A  D  S  Q
TCATTTACGACCTTCAGGCAATGCCCCTGAGTTGCGTTGCTATGCGGAGGCTGACTCGCA   11280

  E  D  A  C  N  I  V  E  T  V  L  S  N  I  K  S  K  L  G  R
AGAGGATGCATGTAATATTGTTGAAACTGTTCTCTCTAATATCAAAAGCAAACTGGGTAG   11340

  End of manB
  A  *
AGCTTAATGCTGTTGATAATAGAGCGTTTCTTTCCAGTAATACTTTGTCTGGTTATCTGG   11400

                    Start of wbaP
                    M  D  R  F  D  N  K  Y  N  P  N  L
TACCCAAGTTGAGGGTGAGAATTAAATGGATCGTTTTGATAATAAGTATAACCCAAATTT   11460

  C  K  I  L  L  A  I  S  D  L  L  F  F  N  V  A  L  W  A  S
ATGCAAAATATTATTGGCTATATCAGATTTACTGTTTTTTAATGTAGCCTTATGGGCATC   11520
```

Figure 9/10

```
  L  G  V  V  Y  L  I  F  D  E  V  Q  R  F  V  P  Q  E  Q  L
GTTAGGAGTTGTATATTTAATCTTTGATGAAGTTCAGCGATTTGTACCACAAGAGCAATT  11580

  D  N  R  F  I  S  H  F  I  L  S  I  V  C  V  G  W  F  W  V
AGATAATCGATTTATATCACATTTTATTCTATCTATAGTATGCGTTGGATGGTTTTGGGT  11640

  R  L  R  H  Y  T  Y  R  K  P  F  W  Y  E  L  K  E  V  I  R
TCGACTGCGTCACTATACATATCGAAAGCCATTCTGGTATGAGTTGAAAGAGGTTATTCG  11700

  T  I  V  I  F  A  V  F  D  L  A  L  I  A  F  T  K  W  Q  F
TACTATCGTTATTTTTGCTGTGTTTGATTTGGCTTTAATTGCGTTTACAAAATGGCAGTT  11760

  S  R  Y  V  W  V  F  C  W  T  F  A  I  I  L  V  P  F  F  R
TTCACGCTATGTCTGGGTGTTTTGTTGGACTTTTGCCATAATCCTGGTGCCTTTTTTTCG  11820

  A  L  T  K  H  L  L  N  K  L  G  I  W  K  K  K  T  I  I  L
CGCACTTACAAAGCATTTATTGAACAAGCTAGGTATCTGGAAGAAAAAAACTATCATCCT  11880

  G  S  G  Q  N  A  R  G  A  Y  S  A  L  Q  S  E  E  M  M  G
TGGGAGCGGACAGAATGCTCGTGGTGCATATTCTGCGCTGCAAAGTGAGGAGATGATGGG  11940

  F  D  V  I  A  F  F  D  T  D  A  S  D  A  E  I  N  M  L  P
GTTTGATGTTATCGCTTTTTTTGATACGGATGCGTCAGATGCTGAAATAAATATGTTGCC  12000

  V  I  K  D  T  E  T  I  W  D  L  N  R  T  G  D  V  H  Y  I
GGTGATAAAGGACACTGAGACTATTTGGGATTTAAATCGTACAGGTGATGTCCATTATAT  12060

  L  A  Y  E  Y  T  E  L  E  K  T  H  F  W  L  R  E  L  S  K
CCTTGCTTATGAATACACCGAGTTGGAGAAAACACATTTTTGGCTACGTGAACTTTCAAA  12120

  H  H  C  R  S  V  T  V  V  P  S  F  R  G  L  P  L  Y  N  T
ACATCATTGTCGTTCTGTTACTGTCGTCCCCTCGTTTAGAGGATTGCCATTATATAATAC  12180

  D  M  S  F  I  F  S  H  E  V  M  L  L  R  I  Q  N  N  L  A
TGATATGTCTTTTATCTTTAGCCATGAAGTTATGTTATTAAGGATACAAAATAACTTGGC  12240

  K  R  S  S  R  F  L  K  R  T  F  D  I  V  C  S  I  M  I  L
TAAAAGGTCGTCCCGTTTTCTCAAACGGACATTTGATATTGTTTGTTCAATAATGATTCT  12300

  I  I  A  S  P  L  M  I  Y  L  W  Y  K  V  T  R  D  G  G  P
TATAATTGCATCACCACTTATGATTTATCTGTGGTATAAAGTTACTCGAGATGGTGGTCC  12360

  A  I  Y  G  H  Q  R  V  G  R  H  G  K  L  F  P  C  Y  K  F
GGCTATTTATGGTCACCAGCGAGTAGGTCGGCATGGAAAACTTTTTCCATGCTACAAATT  12420

  R  S  M  V  M  N  S
TCGTTCTATGGTTATGAATTC  12441
```

Figure 9/11

GAATTCGGGAGGCGCAATGAAAGTCAGCTTTTTTCTGCTGAAATTTCCACTCTCATCGGA 60

AACCTTTGTGCTGAATCAGATTACTGCGTTTATTGATATGGGCCATGAGGTGGAGATTGT 120

CGCGTTACAAAAAGGCGATACCCAACATACTCACGCCGCCTGGGAGAAGTATGGCCTGGC 180

GGCGAAAACCCGCTGGTTACAGGATGAGCCCCAGGGACGGCTGGCGAAACTGCGCTACCG 240

GGCATGTAAAACGCTGCCGGGGCTGCATCGGGCGGCGACCTGGAAAGCGCTCAATTTTAC 300

CCGCTATGGCGATGAATCACGCAATTTGATCCTTTCCGCGATTTGCGCGCAGGTGAGCCA 360

GCCTTTTGTGGCGGATGTGTTTATCGCACACTTTGGTCCGGCGGGCGTGACGGCGGCCAA 420

ACTACGCGAACTGGGCGTGCTTCGCGGCAAAATCGCGACTATTTTCCACGGGATTGATAT 480

CTCTAGTCGTGAGGTGCTCAGTCATTACACGCCGGAGTATCAGCAGTTGTTTCGTCGTGG 540

CGATCTGATGCTGCCCATCAGCGATCTGTGGGCCGGTCGCCTGAAAAGTATGGGCTGTCC 600

GCCGGAAAAGATTGCCGTTTCGCGCATGGGCGTCGACATGACGCGTTTTACCCATCGTTC 660

GGTGAAAGCGCCAGGGATGCCGCTGGAGATGATTTCCGTCGCGCGCCTGACAGAAAAAAA 720

AGGCCTGCATGTGGCGATTGAAGCCTGTCGGCAACTGAAAGCACAGGGCGTGGCGTTTCG 780

CTACCGCATTCTGGGGATTGGCCCGTGGGAACGTCGGCTGCGCACGCTCATCGAGCAGTA 840

TCAGCTAGAGGATGTCATTGAGATGCCGGGGTTTAAACCGAGCCATGAAGTGAAGGCGAT 900

GCTGGATGACGCCGATGTTTTTTTGCTGCCGTCGATTACCGGTACGGATGGCGATATGGA 960

AGGTATTCCGGTAGCGCTGATGGAGGCGATGGCGGTAGGGATTCCCGTGGTATCTACCGT 1020

GCATAGCGGTATTCCGGAACTGGTGGAGGCCGGCAAATCCGGCTGGCTGGTGCCGGAAAA 1080

CGATGCGCAGGCGCTGGCGGCCCGACTCGCTGAGTTCAGCCGGATTGACCACGACACGCT 1140

GGAGTCGGTGATCACGCGCGCCCGTGAAAAAGTGGCGCAAGATTTTAATCAGCAGGCGAT 1200

TAATCGCCAGTTAGCCAGCCTGCTACAAACGATATAAACGAGGTGGTATGCCCGCGACTA 1260

AATTCTCCCGACGTACCCTCCTGACGGCAGGTTCTGCGCTTGCTGTTCTTCCTTTTCTGC 1320

GCGCCTTGCCGGTACAGGCGCGTGAACCTCGCGAGACCGTCGATATTAAGGATTATCCGG 1380

CGGATGACGGTATCGCCTCGTTCAAACAGGCCTTCGCCGACGGACAGACCGTGGTCGTAC 1440

CGCCAGGATGGGTGTGTGAAAATATCAATGCGGCGATAACGATTCCGGCGGGAAAAACGC 1500

TGCGGGTACAGGGCGCGGTGCGTGGGAATGGCCGGGGACGGTTTATTTTGCAGGACGGGT 1560

GTCAGGTGGTGGGGGAGCAGGGCGGCAGTCTGCACAATGTGACGCTGGATGTTCGCGGGT 1620

CGGACTGTGTGATTAAAGGCGTGGCGATGAGCGGCTTTGGCCCCGTCGCGCAAATTTTCA 1680

TCGGTGGTAAGGAACCGCAGGTGATGCGTAATCTCATTATCGATGACATCACCGTTACCC 1740

ACGCCAACTACGCCATTCTCCGCCAGGGATTTCATAACCAAATGGATGGCGCGCGGATTA 1800

CGCATAGCCGCTTTAGCGATTTACAGGGGGACGCCATTGAGTGGAATGTCGCGATTCACG 1860

ACCGCGACATCCTGATTTCCGATCATGTCATCGAACGCATTAATTGTACCAATGGCAAAA 1920

TCAACTGGGGGATCGGCATCGGGCTGGCGGGTAGCACCTATGACAACAGTTATCCTGAAG 1980

Figure 10/1

ACCAGGCAGTAAAAAACTTTGTGGTGGCCAATATTACCGGATCTGATTGCCGACAGCTTG 2040
TGCACGTAGAAAATGGCAAACATTTCGTCATTCGCAATGTCAAAGCCAAAAACATCACGC 2100
CCGGTTTCAGTAAAAATGCGGGTATTGATAACGCAACGATCGCAATTTATGGCTGTGATA 2160
ATTTCGTCATTGATAATATTGATATGACGAATAGTGCCGGGATGCTCATCGGCTATGGCG 2220
TCGTTAAAGGAAAATACCTGTCAATTCCGCAAAACTTTAAATTAAACGCTATTCGGTTGG 2280
ATAATCGCCAGGTTGCTTATAAATTACGCGGCATTCAAATTTCCTCCGGCAACACCCCCT 2340
CTTTTGTCGCCATCACCAATGTACGGATGACGCGTGCTACGCTGGAACTGCATAATCAAC 2400
CGCAGCACCTCTTTCTGCGCAATATCAACGTGATGCAAACTTCAGCGATTGGCCCGGCGT 2460
TAAAAATGCATTTCGATTTGCGTAAAGATGTACGTGGTCAATTTATGGCCCGCCAGGACA 2520
CGCTGCTTTCCCTCGCTAATGTTCATGCCATCAATGAAAACGGGCAGAGTTCCGTGGATA 2580
TCGACAGGATTAATCACCAAACCGTGAATGTCGAAGCAGTGAATTTTTCGCTGCCGAAGC 2640
GGGGAGGGTAAGTACCGCTATTTTTACGAAAATTCCTGGGAAAAAGTTGTTCATACTTAA 2700
TGTTATGGTGCCGACTAAGACGTAATGTAGAGCGTGCCATCATTATCCCTGGCAGCAGAG 2760
TAATTCATGCTGGCGAAAACAAGCTAAAGAGCTATAATTCAGCAACCATTTTACAGGTGG 2820
AAGAAACAATGATGAATTTGAAAGCAGTTATACCGGTAGCGGGTTTGGGTATGCATATGT 2880
TGCCTGCCACCAAGGCAATCCCAAAAGAGATGCTACCGATCGTCGACAAGCCAATGATTC 2940
AGTACATTGTCGATGAGATTGTGGCTGCAGGGATCAAAGAAATCGTGCTGGTGACTCACG 3000
CGTCTAAAAACGCCGTTGAGAACCACTTCGACACCTCTTATGAACTTGAATCACTTCTTG 3060
AGCAGCGCGTTAAGCGTCAGCTTTTGGCGGAAGTGCAATCTATCTGCCCACCGGGCGTGA 3120
CGATTATGAACGTTCGCCAGGCGCAGCCGTTAGGGCTGGGGCATTCTATTCTGTGCGCGC 3180
GTCCGGTCGTGGGCGATAACCCTTTCATTGTGGTACTCCCGGATATTATTATCGATGATG 3240
CTACCGCCGATCCGCTGCGCTATAACCTTGCGGCGATGGTGGCGCGTTTCAATGAAACGG 3300
GTCGCAGCCAGGTGCTGGCGAAGCGCATGAAAGGTGATTTATCGGAGTATTCCGTTATCC 3360
AGACGAAAGAACCTCTGGATAATGAAGGCAAAGTCAGCCGGATTGTGGAGTTTATCGAAA 3420
AACCGGATCAGCCGCAGACGCTGGATTCCGATTTGATGGCGGTAGGCCGTTATGTGCTTT 3480
CAGCCGACATCTGGGCGGAACTGGAAAGAACCGAACCGGGCGCCTGGGGCCGCATCCAGC 3540
TCACCGATGCCATTGCTGAACTGGCGAAAAAACAGTCGGTTGACGCGATGCTAATGACGG 3600
GTGACAGCTATGACTGCGGTAAAAAAATGGGCTACATGCAGGCATTTGTGAAGTACGGGC 3660
TGCGCAACCTGAAAGAAGGAGCCAAGTTCCGTAAGAGCATAGAGCAGCTTTTGCATGAAT 3720
AAGTATTAACAACCGTGATAAATGGTTGGTGATAAACATAATAACGGCAGTGAACATTCG 3780
AAGCGGCAAGTTGGCTGAAACGAGTGTTGACTGCCGTTTTAGTTTTGTATAAAGGGCTTA 3840
AGTAACAAGGGGTTATCTGGAGCATTTTAATGCTGATTTTATAAGATTAATCCTTGTTTC 3900
CGGATGCAATTAATAAGACAATTAGCGTTTAAGTTTTAGTGAGCTTTGCCCTGCTGGGCG 3960

Figure 10/2

```
AGGTTTGCAACAAGTCGATATGTACGCAGTGCACTGGTAGCTGATGAGCCAGGGGCGGTA      4020

GCGTGTGTAACGACTTGAGCAATTAATTTTTATTGGCAAATTAAATACCACATTAAATAC      4080

                Start of rmlB
                V  K  I  L  I  T  G  G  A  G  F  I  G  S
GCCTTATGGAATAGAAAAGTGAAGATACTTATTACTGGCGGGGCAGGTTTTATTGGATCA      4140

A  V  V  R  H  I  I  K  N  T  Q  D  T  V  V  N  I  D  K  L
GCTGTTGTCCGCCATATTATTAAGAATACACAGGACACTGTAGTTAATATTGATAAATTA      4200

T  Y  A  G  N  L  E  S  L  S  D  I  S  E  S  N  R  Y  N  F
ACCTACGCCGGTAATCTTGAATCCCTTTCTGATATTTCTGAAAGTAATCGCTACAATTTT      4260

E  H  A  D  I  C  D  S  A  E  I  T  R  I  F  E  Q  Y  Q  P
GAACACGCGGATATTTGTGATTCCGCTGAAATAACGCGTATTTTTGAGCAGTACCAGCCG      4320

D  A  V  M  H  L  A  A  E  S  H  V  D  R  S  I  T  G  P  A
GACGCGGTGATGCATTTGGCTGCGGAAAGTCATGTGGACCGTTCGATTACCGGGCCAGCA      4380

A  F  I  E  T  N  I  V  G  T  Y  A  L  L  E  V  A  R  K  Y
GCATTTATTGAAACCAATATCGTCGGCACCTATGCACTTCTTGAAGTTGCGCGTAAATAC      4440

W  S  A  L  G  E  D  K  K  N  N  F  R  F  H  H  I  S  T  D
TGGTCTGCCCTTGGCGAAGATAAAAAAAATAATTTTCGTTTTCATCATATTTCCACTGAT      4500

E  V  Y  G  D  L  P  H  P  D  E  V  E  N  S  V  T  L  P  L
GAAGTTTACGGCGATTTACCGCATCCTGATGAAGTTGAAAACAGCGTTACGCTGCCGTTA      4560

F  T  E  T  T  A  Y  A  P  S  S  P  Y  S  A  S  K  A  S  S
TTTACTGAAACGACGGCATATGCGCCAAGTAGCCCCTATTCTGCGTCAAAAGCATCCAGC      4620

D  H  L  V  R  A  W  R  R  T  Y  G  L  P  T  I  V  T  N  C
GATCATTTAGTCCGTGCCTGGCGGCGTACCTATGGTCTACCAACGATCGTTACCAATTGT      4680

S  N  N  Y  G  P  Y  H  F  P  E  K  L  I  P  L  V  I  L  N
TCTAATAACTATGGCCCTTATCACTTCCCTGAAAAACTGATTCCGTTGGTCATTTTGAAC      4740

A  L  E  G  K  P  L  P  I  Y  G  K  G  D  Q  I  R  D  W  L
GCACTGGAAGGAAAGCCTTTGCCAATTTATGGCAAAGGGGATCAGATTCGCGATTGGCTA      4800

Y  V  E  D  H  A  R  A  L  H  M  V  V  T  E  G  K  A  G  E
TATGTAGAAGATCATGCTCGCGCGCTTCATATGGTAGTGACTGAAGGCAAGGCAGGGGAG      4860

T  Y  N  I  G  G  H  N  E  K  K  N  L  D  V  V  F  T  I  C
ACTTATAACATTGGTGGACACAATGAGAAGAAAAATCTCGATGTGGTATTTACCATCTGT      4920

D  L  L  D  E  I  V  P  K  A  T  S  Y  R  E  Q  I  T  Y  V
GATCTGCTGGATGAGATTGTACCCAAAGCGACTTCTTATCGTGAACAAATCACTTATGTC      4980

A  D  R  P  G  H  D  R  R  Y  A  I  D  A  G  K  I  S  R  E
GCGGATCGTCCGGGCCATGATCGTCGTTATGCCATTGATGCAGGTAAAATTAGCCGCGAA      5040

L  G  W  K  P  L  E  T  F  E  S  G  I  R  K  T  V  E  W  Y

TTAGGCTGGAAACCGCTGGAGACCTTTGAAAGCGGTATTCGTAAAACAGTGGAATGGTAC      5100

L  A  N  T  Q  W  V  N  N  V  K  S  G  A  Y  Q  S  W  I  E
CTTGCAAATACTCAATGGGTAAACAATGTTAAAAGTGGGGCGTATCAGAGTTGGATAGAA      5160

End of rmlB    Start of rmlD
Q  N  Y  E  G  R  Q  *
                        M  N  I  L  L  F  G  K  T  G  Q  V
CAGAACTATGAAGGACGCCAGTAATGAATATCTTACTTTTTGGTAAGACAGGGCAAGTAG      5220
```

Figure 10/3

```
G  W  E  L  Q  R  S  L  A  P  V  G  N  L  I  A  L  D  V  H
GCTGGGAGTTGCAACGTTCTCTGGCACCGGTAGGGAATCTGATTGCCCTGGATGTCCATT    5280

S  K  E  F  C  G  D  F  S  N  P  K  G  V  A  E  T  V  R  K
CAAAAGAGTTTTGCGGTGATTTTAGTAATCCGAAAGGCGTTGCCGAAACCGTTCGTAAGC    5340

L  R  P  D  V  I  V  N  A  A  A  H  T  A  V  D  K  A  E  S
TTCGTCCCGATGTGATTGTTAACGCAGCAGCCCATACTGCAGTAGATAAAGCAGAGTCTG    5400
                                *
E  P  E  L  A  Q  L  L  N  A  T  S  V  E  A  I  A  K  A  A
AACCAGAACTGGCGCAGTTACTTAACGCCACCAGTGTGGAAGCCATCGCTAAAGCAGCCA    5460

N  E  T  G  A  W  V  V  H  Y  S  T  D  Y  V  F  P  G  T  G
ACGAAACTGGCGCATGGGTAGTGCATTATTCAACCGATTATGTATTTCCTGGTACCGGCG    5520

D  I  P  W  Q  E  T  D  A  T  S  P  L  N  V  Y  G  K  T  K
ATATCCCATGGCAGGAAACGGACGCTACGTCGCCGCTGAATGTCTATGGCAAAACCAAAC    5580

L  A  G  E  K  A  L  Q  D  N  C  P  K  H  L  I  F  R  T  S
TGGCGGGAGAAAAGGCCCTGCAGGATAACTGCCCTAAACACCTTATCTTCCGCACCAGTT    5640

W  V  Y  A  G  K  G  N  N  F  A  K  T  M  L  R  L  A  K  E
GGGTTTATGCAGGTAAGGGCAATAATTTCGCAAAGACAATGCTTCGTCTGGCGAAAGAGC    5700

R  Q  T  L  S  V  I  N  D  Q  Y  G  A  P  T  G  A  E  L  L
GTCAGACACTTTCAGTCATTAACGATCAGTACGGTGCGCCAACCGGTGCGGAATTACTGG    5760

A  D  C  T  A  H  A  I  R  V  A  L  N  K  P  E  V  A  G  L
CTGACTGTACGGCGCATGCGATCCGTGTGGCGTTAAATAAACCAGAAGTCGCAGGTCTTT    5820

Y  H  L  V  A  G  G  T  T  T  W  H  D  Y  A  A  L  V  F  D
ACCATCTGGTTGCCGGGGGAACCACAACCTGGCATGACTACGCGGCCTTAGTCTTTGACG    5880

E  A  R  K  A  G  I  T  L  A  L  T  E  L  N  A  V  P  T  S
AGGCGCGCAAAGCAGGATAACGCTTGCGCTGACTGAGCTTAATGCTGTGCCGACCAGCG    5940

A  Y  P  T  P  A  S  R  P  G  N  S  R  L  N  T  E  K  F  Q
CCTACCCGACGCCGGCGAGCAGACCAGGCAATTCGCGTCTCAATACTGAAAAGTTTCAGC    6000

R  N  F  D  L  I  L  P  Q  W  E  L  G  V  K  R  M  L  T  E
GTAATTTTGACCTTATTCTGCCTCAATGGGAATTAGGAGTTAAGCGTATGCTGACTGAAA    6060

                    End of rmlD
M  F  T  T  T  T  I  *
TGTTTACGACGACAACCATC TAATAAATTTAAATGCCCATCAGGGCATTTTCTATGAATG    6120

          Start of rmlA
             M  K  T  R  K  G  I  I  L  A  G  G  S  G  T  R  L
AGAAATGGAAATGAAAACGCGTAAGGGCATTATTTTAGCGGGGGGCTCCGGCACCCGTCT    6180

  Y  P  V  T  M  A  V  S  K  Q  L  L  P  I  Y  D  K  P  M  I
TTATCCGGTGACCATGGCGGTAAGTAAGCAATTGCTACCAATTTATGATAAACCGATGAT    6240

  Y  Y  P  L  S  T  L  M  L  A  G  I  R  D  I  L  I  I  S  T
TTACTATCCCCTTTCCACGCTTATGCTGGCAGGCATTCGGGATATCCTGATCATCAGTAC    6300

  P  Q  D  T  P  R  F  Q  Q  L  L  G  D  G  S  Q  W  G  L  N
GCCACAGGACACGCCGCGTTTCAACAACTGCTGGGAGACGGCAGCCAGTGGGGGCTGAA    6360

  L  Q  Y  K  V  Q  P  S  P  D  G  L  A  Q  A  F  I  I  G  E
TCTTCAATATAAAGTACAGCCAAGCCCGGATGGCTTAGCACAGGCGTTTATTATTGGTGA    6420

  E  F  I  G  H  D  D  C  A  L  V  L  G  D  N  I  F  Y  G  H
AGAGTTCATTGGTCATGATGATTGTGCATTAGTGCTGGGTGACAATATCTTCTATGGTCA    6480
```


Figure 10/4

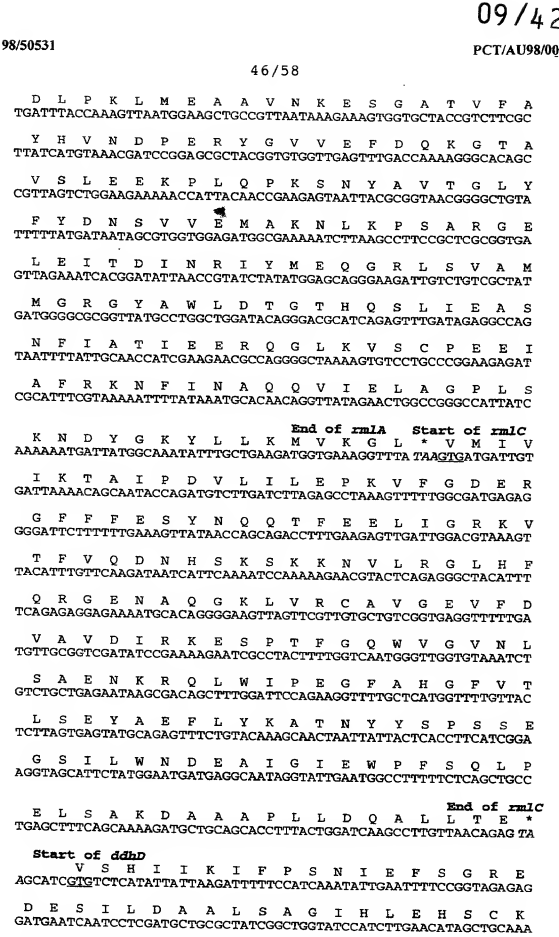
D L P K L M E A A V N K E S G A T V F A
TGATTTACCAAAGTTAATGGAAGCTGCCGTTAATAAAGAAAGTGGTGCTACCGTCTTCGC 6540
Y H V N D P E R Y G V V E F D Q K G T A
TTATCATGTAAACGATCCGGAGCGCTACGGTGTGGTTGAGTTTGACCAAAAGGGCACAGC 6600
V S L E E K P L Q P K S N Y A V T G L Y
CGTTAGTCTGGAAGAAAAACCATTACAACCGAAGAGTAATTACGCGGTAACGGGGCTGTA 6660
F Y D N S V V E M A K N L K P S A R G E
TTTTTATGATAATAGCGTGGTGGAGATGGCGAAAAATCTTAAGCCTTCCGCTCGCGGTGA 6720
L E I T D I N R I Y M E Q G R L S V A M
GTTAGAAATCACGGATATTAACCGTATCTATATGGAGCAGGGAAGATTGTCTGTCGCTAT 6780
M G R G Y A W L D T G T H Q S L I E A S
GATGGGGCGCGGTTATGCCTGGCTGGATACAGGGACGCATCAGAGTTTGATAGAGGCCAG 6840
N F I A T I E E R Q G L K V S C P E E I
TAATTTTATTGCAACCATCGAAGAACGCCAGGGGCTAAAAGTGTCCTGCCCGGAAGAGAT 6900
A F R K N F I N A Q Q V I E L A G P L S
CGCATTTCGTAAAAATTTTATAAATGCACAACAGGTTATAGAACTGGCCGGGCCATTATC 6960
End of rmlA Start of rmlC
K N D Y G K Y L L K M V K G L * V M I V
AAAAAATGATTATGGCAAATATTTGCTGAAGATGGTGAAAGGTTTA TAAGTGATGATTGT 7020
I K T A I P D V L I L E P K V F G D E R
GATTAAAACAGCAATACCAGATGTCTTGATCTTAGAGCCTAAAGTTTTTGGCGATGAGAG 7080
G F F F E S Y N Q Q T F E E L I G R K V
GGGATTCTTTTTTGAAAGTTATAACCAGCAGACCTTTGAAGAGTTGATTGGACGTAAAGT 7140
T F V Q D N H S K S K K N V L R G L H F
TACATTTGTTCAAGATAATCATTCAAAATCCAAAAAGAACGTACTCAGAGGGCTACATTT 7200
Q R G E N A Q G K L V R C A V G E V F D
TCAGAGAGGAGAAAATGCACAGGGGAAGTTAGTTCGTTGTGCTGTCGGTGAGGTTTTTGA 7260
V A V D I R K E S P T F G Q W V G V N L
TGTTGCGGTCGATATCCGAAAAGAATCGCCTACTTTTGGTCAATGGGTTGGTGTAAATCT 7320
S A E N K R Q L W I P E G F A H G F V T
GTCTGCTGAGAATAAGCGACAGCTTTGGATTCCAGAAGGTTTTGCTCATGGTTTTGTTAC 7380
L S E Y A E F L Y K A T N Y Y S P S S E
TCTTAGTGAGTATGCAGAGTTTCTGTACAAAGCAACTAATTATTACTCACCTTCATCGGA 7440
G S I L W N D E A I G I E W P F S Q L P
AGGTAGCATTCTATGGAATGATGAGGCAATAGGTATTGAATGGCCTTTTTCTCAGCTGCC 7500
End of rmlC
E L S A K D A A A P L L D Q A L L T E *
TGAGCTTTCAGCAAAAGATGCTGCAGCACCTTTACTGGATCAAGCCTTGTTAACAGAG TA 7560
Start of ddhD
V S H I I K I F P S N I E F S G R E
AGCATCGTGTCTCATATTATTAAGATTTTTCCATCAAATATTGAATTTTCCGGTAGAGAG 7620
D E S I L D A A L S A G I H L E H S C K
GATGAATCAATCCTCGATGCTGCGCTATCGGCTGGTATCCATCTTGAACATAGCTGCAAA 7680
A G D C G I C E S D L L A G E V V D S K
GCGGGTGATTGTGGTATCTGTGAGTCCGATTTGTTGGCGGGAGAAGTTGTTGACTCCAAA 7740

Figure 10/5

```
G N I F G Q G D K I L T C C C K P K T A
GGTAATATTTTTGGACAGGGTGATAAAATACTAACCTGCTGCTGTAAACCTAAAACCGCC   7800

L E L N A H F F P E L A G Q T K K I V P
CTTGAGCTAAATGCCGCATTTTTTTCCTGAACTAGCTGGACAGACAAAAAAAATTGTCCCA   7860

C K V N S A V L V S G D V M T L K L R T
TGCAAGGTAAATAGTGCTGTACTGGTTTCAGGCGATGTTATGACTTTGAAGTTACGCACA   7920

P P T A K I G F L P G Q Y I N L H Y K G
CCACCAACAGCAAAAATTGGCTTCCTTCCAGGGCAGTATATCAATTTACATTATAAAGGT   7980

V T R S Y S I A N S D E S N G I E L H V
GTAACTCGCAGTTATTCTATCGCTAATAGTGATGAGTCGAATGGTATTGAGTTGCATGTA   8040

R N V P N G Q M S S L I F G E L Q E N T
AGGAATGTTCCCAATGGTCAGATGAGTTCGCTCATTTTTGGGGAGTTACAAGAAAATACT   8100

L M R I E G P C G T F F I R E S D R P I
CTTATGCGCATTGAAGGGCCTTGCGGAACATTTTTTATTCGTGAAAGTGACAGACCTATA   8160

I F L A G G T G F A P V K S M V E H L I
ATCTTCCTTGCAGGCGGTACTGGATTCGCTCCAGTTAAATCAATGGTTGAGCATCTCATT   8220

Q G K C R R E I Y I Y W G M Q Y S K D F
CAGGGAAAATGTCGTCGTGAGATCTACATTTACTGGGGAATGCAATATAGTAAAGATTTT   8280

Y S A L P Q Q W S E Q H D N V H Y I P V
TACTCTGCATTACCGCAGCAGTGGAGTGAACAGCACGACAACGTTCATTATATCCCTGTT   8340

V S G D D A E W G G R K G F V H H A V M
GTTTCTGGTGATGACGCCGAATGGGGGGGAAGAAAGGGATTTGTCCATCATGCCGTGATG   8400

D D F D S L E F F D I Y A C G S P V M I
GATGATTTTGATTCTCTAGAGTTCTTCGATATATATGCATGTGGTTCACCTGTGATGATC   8460

D A S K K D F M M K N L S V E H F Y S D
GATGCCAGTAAAAAGGACTTTATGATGAAAAATCTCTCTGTAGAACATTTCTATTCTGAT   8520

                    End of ddhD   Start of ddhA
A F T A S N N I E D N L *
                                  M K A V I L A G
GCATTTACCGCATCTAATAATATTGAGGATAATTTATATGAAAGCGGTCATCCTGGCTGGTG   8580

G L G T R L S E E T I V K P K P M V E I
GACTTGGTACCAGACTAAGTGAAGAAACAATTGTAAAACCAAAACCGATGGTAGAAATTG   8640

G G K P I L W H I M K M Y S V H G I K D
GTGGCAAGCCTATTCTTTGGCACATTATGAAAATGTATTCTGTGCATGGTATCAAGGATT   8700

F I I C C G Y K G Y V I K E Y F A N Y F
TTATTATCTGCTGTGGTTATAAAGGATATGTGATTAAAGAATATTTTGCGAACTACTTCC   8760

L H M S D V T F H M A E N R M E V H H K
TTCACATGTCAGATGTAACATTCCATATGGCTGAAAACCGTATGGAAGTTCACCATAAAC   8820

R V E P W N V T L V D T G D S S M T G G
GTGTTGAACCATGGAATGTCACATTGGTTGATACGGGTGATTCTTCAATGACTGGTGGTC   8880

R L K R V A E Y V K D D E A F L F T Y G
GTCTGAAACGTGTTGCTGAATACGTAAAAGATGACGAGGCTTTCCTGTTTACTTATGGTG   8940

D G V A D L D I K A T I D F H K A H G K
ATGGCGTTGCCGACCTTGATATCAAAGCGACTATCGATTTCCATAAGGCTCACGGTAAGA   9000
```


Figure 10/6

```
K  A  T  L  T  A  T  F  P  P  G  R  F  G  A  L  D  I  R  A
AAGCGACTTTAACAGCTACTTTTCCACCAGGACGCTTTGGCGCATTAGATATCCGAGCTG                9060

G  Q  V  R  S  F  Q  E  K  P  K  G  D  G  A  M  I  N  G  G
GTCAGGTCCGGTCATTCCAGGAAAAACCGAAAGGCGATGGGGCAATGATCAATGGTGGTT                9120

F  F  V  L  N  P  S  V  I  D  L  I  D  N  D  A  T  T  W  E
TCTTTGTGTTGAATCCATCGGTTATCGATCTCATCGATAACGATGCAACAACCTGGGAAC                9180

Q  E  P  L  M  T  L  A  Q  Q  G  E  L  M  A  F  E  H  P  G
AAGAGCCATTAATGACATTGGCACAACAGGGGGAGTTAATGGCTTTTGAACACCCAGGTT                9240

F  W  Q  P  M  D  T  L  R  D  K  V  Y  L  E  G  L  W  E  K
TCTGGCAGCCGATGGATACCCTACGTGATAAAGTTTACCTCGAAGGGCTGTGGGAAAAAG                9300

                End of ddhA  Start of ddhB
                                M  I  D  K  N  F  W  Q  G
G  K  A  P  W  K  T  W  E  *
GTAAAGCTCCGTGGAAAACCTGGGAGTAACTAGATGATTGATAAAAATTTTTGGCAAGGT                9360

 K  R  V  F  V  T  G  H  T  G  F  K  G  S  W  L  S  L  W  L
AAACGTGTATTCGTTACCGGCCATACTGGCTTTAAAGGAAGCTGGCTTTCGCTATGGCTG                9420

 T  E  M  G  A  I  V  K  G  Y  A  L  D  A  P  T  V  P  S  L
ACTGAAATGGGTGCAATTGTAAAAGGCTATGCACTTGATGCGCCAACTGTTCCAAGTTTA                9480

 F  E  I  V  R  L  N  D  L  M  E  S  H  I  G  D  I  R  D  F
TTTGAGATAGTGCGTCTTAATGATCTTATGGAATCTCATATTGGCGACATTCGTGATTTT                9540

 E  K  L  R  N  S  I  A  E  F  K  P  E  I  V  F  H  M  A  A
GAAAAGCTGCGCAATTCTATTGCAGAATTTAAGCCAGAAATTGTTTTCCATATGGCAGCC                9600

 Q  P  L  V  R  L  S  Y  E  Q  P  I  E  T  Y  S  T  N  V  M
CAGCCTTTAGTGCGCCTATCTTATGAACAGCCAATCGAAACATACTCAACAAATGTTATG                9660

 G  T  V  H  L  L  E  T  V  K  Q  V  G  N  I  K  A  V  V  N
GGTACTGTCCATTTGCTTGAAACAGTTAAGCAAGTAGGTAACATAAAGGCAGTCGTAAAT                9720

 I  T  S  D  K  C  Y  D  N  R  E  W  V  W  G  Y  R  E  N  E
ATCACCAGTGATAAGTGCTACGACAATCGTGAGTGGGTGTGGGGCTATCGTGAGAACGAA                9780

 P  M  G  G  Y  D  P  Y  S  N  S  K  G  C  A  E  L  V  A  S
CCCATGGGAGGGTACGATCCATACTCTAATAGTAAAGGTTGTGCAGAATTAGTCGCGTCT                9840

 A  F  R  N  S  F  F  N  P  A  N  Y  E  Q  H  G  V  G  L  A
GCATTCCGGAACTCATTCTTCAATCCTGCAAATTATGAGCAACATGGCGTTGGTTTGGCG                9900

 S  V  R  A  G  N  V  I  G  G  G  D  W  A  K  D  R  L  I  P
TCTGTGAGGGCTGGTAATGTCATAGGCGGAGGCGATTGGGCTAAAGACCGTTTAATTCCC                9960

 D  I  L  R  S  F  E  N  N  Q  Q  V  I  I  R  N  P  Y  S  I
GATATTCTGCGCTCATTTGAAAATAACCAGCAGGTTATTATTCGAAACCCATATTCTATC               10020

 R  P  W  Q  H  V  L  E  P  L  S  G  Y  I  V  V  A  Q  R  L
CGTCCCTGGCAGCATGTACTGGAGCCTCTTTCTGGTTACATTGTGGTGGCGCAACGCTTA               10080

 Y  T  E  G  A  K  F  S  E  G  W  N  F  G  P  R  D  E  D  A
TATACAGAAGGTGCTAAGTTTTCTGAAGGATGGAATTTCGGCCCGCGTGATGAAGATGCG               10140

 K  T  V  E  F  I  V  D  K  M  V  T  L  W  G  D  D  A  S  W
AAGACGGTCGAATTTATTGTTGACAAGATGGTCACGCTTTGGGGTGATGATGCAAGCTGG               10200

 L  L  D  G  E  N  H  P  H  E  A  H  Y  L  K  L  D  C  S  K
TTACTGGATGGTGAGAATCATCCTCATGAGGCACATTACCTGAAACTGGATTGCTCTAAA               10260
```

Figure 10/7

```
 A N M Q L G W H P R W G L T E T L G R I
GCAAATATGCAATTAGGATGGCATCCGCGTTGGGGATTGACTGAAACACTTGGTCGCATC  10320

 V K W H K A W I R G E D M L I C S K R E
GTAAAATGGCATAAAGCATGGATTCGCGGCGAAGATATGTTGATTTGTTCAAAGCGTGAA  10380

                End of ddhB
 I S D Y M S A T T R *
ATCAGCGACTATATGTCTGCAACTACTCGTTAAGAAAATAAGTTTAAGGAATCAAAGTAA  10440

  Start of ddhC
M T A N N L R E Q I S Q L V A Q Y A N E
TGACAGCAAATAACCTGCGTGAGCAAATCTCTCAGCTTGTCGCTCAGTATGCGAATGAGG  10500

A L S P K P F V A G T S V V P P S G K V
CATTGAGCCCGAAACCTTTTGTTGCAGGTACAAGCGTTGTGCCTCCTTCCGGGAAGGTTA  10560

I G A K E L Q L M V E A S L D G W L T T
TTGGTGCCAAAGAGTTACAATTGATGGTTGAGGCGTCTCTTGATGGATGGCTAACTACTG  10620

G R F N D A F E K K L G E F I G V P H V
GTCGTTTCAATGATGCCTTTGAAAAAAAACTTGGGGAATTTATTGGGGTTCCTCATGTTT  10680

L T T T S G S S A N L L A L T A L T S P
TAACGACAACATCTGGCTCTTCGGCAAACTTGCTGGCACTGACTGCGCTGACTTCCCCAA  10740

K L G E R A L K P G D E V I T V A A G F
AATTAGGCGAGCGAGCTCTCAAACCTGGTGATGAGGTTATTACTGTCGCTGCTGGCTTCC  10800

P T T V N P A I Q N G L I P V F V D V D
CGACTACAGTTAACCCGGCGATCCAGAATGGTTTAATACCGGTATTCGTGGATGTTGATA  10860

I P T Y N I D A S L I E A A V T E K S K
TCCCGACATATAATATCGATGCCTCTCTCATTGAAGCTGCAGTTACTGAGAAATCAAAAG  10920

A I M I A H T L G N A F N L S E V R R I
CGATAATGATCGCTCATACACTCGGTAATGCATTTAACCTGAGTGAAGTTCGTCGGATTG  10980

A D K Y N L W L I E D C C D A L G T T Y
CCGATAAATATAACTTATGGTTGATTGAAGACTGCTGTGATGCCCTTGGGACGACTTATG  11040

E G Q M V G T F G D I G T V S F Y P A H
AAGGCCAGATGGTAGGTACCTTTGGTGACATCGGAACCGTTAGTTTTTATCCGGCTCACC  11100

H I T M G E G G A V F T K S G E L K K I
ATATCACAATGGGTGAAGGCGGTGCTGTATTCACCAAGTCAGGTGAACTGAAGAAAATTA  11160

I E S F R D W G R D C Y C A P G C D N T
TTGAGTCGTTCCGTGACTGGGGCCGGGATTGTTATTGTGCGCCAGGATGCGATAACACCT  11220

C G K R F G Q Q L G S L P Q G Y D H K Y
GCGGTAAACGTTTTGGTCAGCAATTGGGATCACTTCCTCAAGGCTATGATCACAAATATA  11280

T Y S H L G Y N L K I T D M Q A A C G L
CTTATTCCCACCTCGGATATAATCTCAAAATCACGGACATGCAGGCAGCATGTGGTCTGG  11340

A Q L E R V E E F V E Q R K A N F S Y L
CTCAGTTGGAGCGCGTAGAAGAGTTTGTAGAGCAGCGTAAAGCTAACTTTTCCTATCTGA  11400

K Q G L Q S C T E F L E L P E A T E K S
AACAGGGCTTGCAATCTTGCACTGAATTCCTCGAATTACCAGAAGCAACAGAGAAATCAG  11460

D P S W F G F P I T L K E T S G V N R V
ATCCATCCTGGTTTGGCTTCCCTATCACCCTGAAAGAAACTAGCGGTGTTAACCGTGTCG  11520
```

Figure 10/8

```
E  L  V  K  F  L  D  E  A  K  I  G  T  R  L  L  F  A  G  N
AACTGGTGAAATTCCTTGATGAAGCAAAAATCGGTACACGTTTACTGTTTGCTGGAAATC  11580

L  I  R  Q  P  Y  F  A  N  V  K  Y  R  V  V  G  E  L  T  N
TGATTCGCCAACCGTATTTTGCTAATGTGAAATATCGTGTAGTGGGTGAGTTGACAAATA  11640

T  D  R  I  M  N  Q  T  F  W  I  G  I  Y  P  G  L  T  T  E
CCGACCGTATAATGAATCAAACGTTCTGGATTGGTATTTATCCAGGCTTGACTACAGAGC  11700

                                        End of ddbC
H  L  D  Y  V  V  S  K  F  E  E  F  F  G  L  N  F  *
ATTTAGATTATGTAGTTAGCAAGTTTGAAGAGTTCTTTGGTTTGAATTTC TAATTCAATT  11760

          Start of abe
                  M  T  F  L  K  E  Y  V  I  V  S  G  A
TATTCTATCTGGTGATTGCGATGACCTTTTTGAAAGAATATGTAATTGTCAGTGGGGCTT  11820

S  G  F  I  G  K  H  L  L  E  A  L  K  K  S  G  I  S  V  V
CCGGCTTTATTGGTAAGCATTTACTCGAAGCGCTAAAAAAATCGGGGATTTCAGTTGTCG  11880

A  I  T  R  D  V  I  K  N  N  S  N  A  L  A  N  V  R  W  C
CAATCACTCGAGATGTAATAAAAAATAATAGTAATGCATTAGCTAATGTTAGATGGTGCA  11940

S  W  D  N  I  E  L  L  V  E  E  L  S  I  D  S  A  L  I  G
GTTGGGATAATATCGAATTATTAGTCGAGGAGTTATCAATTGATTCTGCATTAATTGGTA  12000

I  I  H  L  A  T  E  Y  G  H  K  T  S  S  L  I  N  I  E  D
TCATTCATTTGGCAACAGAATATGGGCATAAAACATCATCTCTCATAAATATTGAAGATG  12060

A  N  V  I  K  P  L  K  L  L  D  L  A  I  K  Y  R  A  D  I
CAAATGTTATAAAACCATTAAAGCTTCTTGATTTGGCAATAAAATATCGGGCGGATATCT  12120

F  L  N  T  D  S  F  F  A  K  K  D  F  N  Y  Q  H  M  R  P
TTTTAAATACAGATAGTTTTTTTGCCAAGAAAGATTTTAATTATCAACATATGCGGCCTT  12180

Y  I  I  T  K  R  H  F  D  E  I  G  H  Y  Y  A  N  M  H  D
ATATAATTACTAAAAGACACTTTGATGAAATTGGGCATTATTATGCTAATATGCATGACA  12240

I  S  F  V  N  M  R  L  E  H  V  Y  G  P  G  D  G  E  N  K
TTTCATTTGTAAACATGCGATTAGAGCATGTATATGGGCCTGGGGATGGTGAAAATAAAT  12300

F  I  P  Y  I  I  D  C  L  N  K  K  Q  S  C  V  K  C  T  T
TTATTCCATACATTATCGACTGCTTAAATAAAAAACAGAGTTGCGTGAAATGTACAACAG  12360

G  E  Q  I  R  D  F  I  F  V  D  D  V  V  N  A  Y  L  T  I
GCGAACAGATAAGAGACTTTATTTTTGTAGATGATGTGGTAAATGCTTATTTAACTATAT  12420

L  E  N  R  K  E  V  P  S  Y  T  E  Y  Q  V  G  T  G  A  G
TAGAAAATAGAAAAGAAGTACCTTCATATACTGAGTATCAAGTTGGAACTGGTGCTGGGG  12480

V  S  L  K  D  F  L  V  Y  L  Q  N  T  M  M  P  G  S  S  S
TAAGTTTGAAAGATTTTCTGGTTTATTTGCAAAATACTATGATGCCAGGTTCATCGAGTA  12540

I  F  E  F  G  A  I  E  Q  R  D  N  E  I  M  F  S  V  A  N
TATTTGAATTTGGTGCGATAGAGCAAAGAGATAATGAAATAATGTTCTCTGTAGCAAATA  12600

N  K  N  L  K  A  M  G  W  K  P  N  F  D  Y  K  K  G  I  E
ATAAAAATTTAAAAGCAATGGGCTGGAAACCAAATTTCGATTATAAAAAGGAATTGAAG  12660

          End of abe
E  L  L  K  R  L  *
AACTACTGAAACGGTTA TGAGATTTTCATGATCTTTTAATAAATAAATCGTTAACAAATT  12720

                                        Start of wzx
                                              V  K  V  Q  L  L
AGTCGCGTTATGTTGTAAAAACTAAGTCGTTTAATTGCATA GTGAAAGTTCAATTGTTAA  12780
```

Figure 10/9

```
K I P S H L I V A G S S W L S K I I I A
AAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTATCCAAAATAATAATTGCCG   12840

G V Q L A S I S Y L I S M L G E E K Y A
GGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAA  12900

I F S L L T G L L V W C S A V D F G I G
TCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTA  12960

T G L Q N Y I S E C R A K N K S Y D A Y
CAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATA  13020

I K S A L H L S F I A I I F F I A L F Y
TTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATA  13080

I F S G V I S A K Y L S S F H E V L Q D
TTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACA  13140

K T R M L F F T S C L V F S S I G I G A
AAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTA  13200

I A Y K I L F A E L V G W K A N L L N A
TTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCAT  13260

L S Y M I G M L G L L Y I Y Y R G I S V
TATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTG  13320

D I K L S L I V L Y L P V G M I S L C Y
ACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATA  13380

I V Y R Y I K L Y H V K T T K S H Y I A
TTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAA  13440

I L R R S S G F F L F T L L S I V V L Q
TTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAA  13500

T D Y M V I S Q R L T P A D I V Q Y T V
CAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTAA  13560

T M K I F G L V F F I Y T A I L Q A L W
CGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGC  13620

P I C A E L R V K Q Q W K K L N K M I G
CTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTG  13680

V N I L L G S L Y V V G C T I F I Y L F
TCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTA  13740

K E Q I F S V I A K D I N Y Q V S I L S
AAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTT  13800

F M L I G I Y F C I R V W C D T Y A M L
TTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTAT  13860

L Q S M N Y L K I L W I L V P L Q A I I
TGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTACCACTACAAGCAATAATTG  13920

G G I A Q W Y F S S T L G I S G V L L G
GTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATCAGTGGAGTGCTGCTTGGCT  13980

L I I S F A L T V F W G L P L T Y L I K
TGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCACTAACTTACTTAATTAAGG  14040
```

Figure 10/10

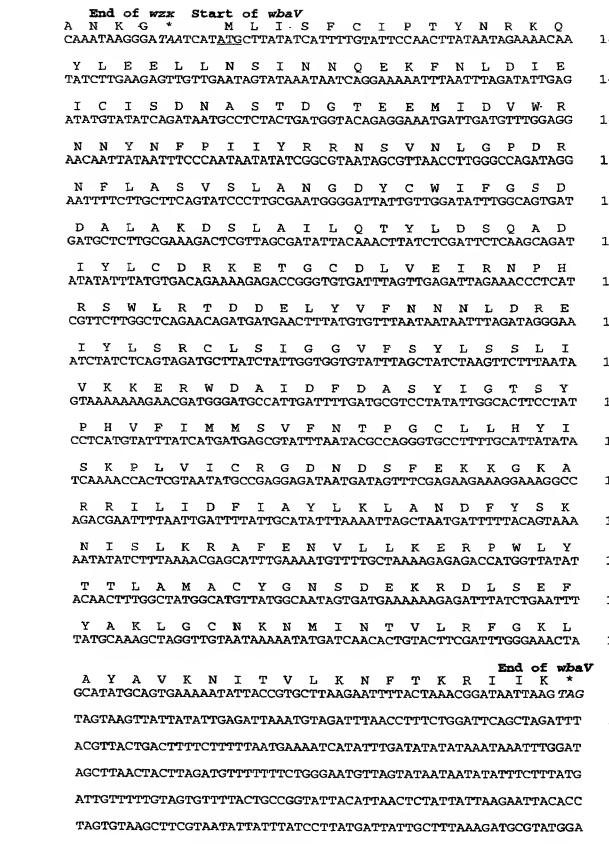
End of wzx Start of wbaV
A N K G * M L I S F C I P T Y N R K Q
CAAATAAGGGATAATCATATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAA 14100
Y L E E L L N S I N N Q E K F N L D I E
TATCTTGAAGAGTTGTTGAATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAG 14160
I C I S D N A S T D G T E E M I D V W R
ATATGTATATCAGATAATGCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGG 14220
N N Y N F P I I Y R R N S V N L G P D R
AACAATTATAATTTCCCAATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGG 14280
N F L A S V S L A N G D Y C W I F G S D
AATTTTCTTGCTTCAGTATCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGAT 14340
D A L A K D S L A I L Q T Y L D S Q A D
GATGCTCTTGCGAAAGACTCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGAT 14400
I Y L C D R K E T G C D L V E I R N P H
ATATATTTATGTGACAGAAAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCAT 14460
R S W L R T D D E L Y V F N N N L D R E
CGTTCTTGGCTCAGAACAGATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAA 14520
I Y L S R C L S I G G V F S Y L S S L I
ATCTATCTCAGTAGATGCTTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATA 14580
V K K E R W D A I D F D A S Y I G T S Y
GTAAAAAAAGAACGATGGGATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTAT 14640
P H V F I M M S V F N T P G C L L H Y I
CCTCATGTATTTATCATGATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATA 14700
S K P L V I C R G D N D S F E K K G K A
TCAAAACCACTCGTAATATGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCC 14760
R R I L I D F I A Y L K L A N D F Y S K
AGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAA 14820
N I S L K R A F E N V L L K E R P W L Y
AATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATAT 14880
T T L A M A C Y G N S D E K R D L S E F
ACAACTTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTT 14940
Y A K L G C N K N M I N T V L R F G K L
TATGCAAAGCTAGGTTGTAATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTA 15000
End of wbaV
A Y A V K N I T V L K N F T K R I I K *
GCATATGCAGTGAAAAATATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAG TAG 15060
TAGTAAGTTATTATATTGAGATTAAATGTAGATTTAACCTTTCTGGATTCAGCTAGATTT 15120
ACGTTACTGACTTTTCTTTTTAATGAAAATCATATTTGATATATATAAATAAATTTGGAT 15180
AGCTTAACTACTTAGATGTTTTTTTCTGGGAATGTTAGTATAATAATATATTTCTTTATG 15240
ATTGTTTTTGTAGTGTTTTACTGCCGGTATTACATTAACTCTATTATTAAGAATTACACC 15300
TAGTGTAAGCTTCGTAATATTATTTATCCTTATGATTATTGCTTTAAAGATGCGTATGGA 15360
Start of wbaU
M I V N L S R L G K S G T G
AAAACGGAGAGCTATTCAATGATCGTAAACCTATCACGTTTAGGTAAAAGTGGTACGGGA 15420

Figure 10/11

```
 M  W  Q  Y  S  I  K  F  L  T  A  L  R  E  I  A  D  V  D  A
ATGTGGCAATACTCGATTAAATTTTTAACGGCACTGCGAGAAATAGCTGATGTTGACGCA   15480

 I  I  C  S  K  V  H  A  D  Y  F  E  K  L  G  Y  A  V  V  T
ATAATCTGTAGCAAGGTACACGCTGATTATTTTGAAAAGCTCGGTTATGCAGTAGTTACT   15540

 V  P  N  I  V  S  N  T  S  K  T  S  R  L  R  P  L  V  W  Y
GTTCCGAATATTGTTAGCAACACATCAAAAACATCGCGACTTAGACCATTAGTATGGTAT   15600

 V  Y  S  Y  W  L  A  L  R  V  L  I  K  F  G  N  K  K  L  V
GTATATAGTTACTGGCTTGCGCTGAGGGTTTTAATTAAGTTGGTAATAAAAAATTGGTG   15660

 C  T  T  H  H  T  I  P  L  L  R  N  Q  T  I  T  V  H  D  I
TGTACTACACATCACACTATCCCCTTACTGAGAAACCAAACGATAACCGTACATGATATA   15720

 R  P  F  Y  Y  P  D  S  F  I  Q  K  V  Y  F  R  F  L  L  K
AGACCTTTTTATTATCCAGATAGTTTTATTCAGAAAGTGTATTTTCGCTTTTTATTAAAA   15780

 M  S  V  K  R  C  K  H  V  L  T  V  S  Y  T  V  K  D  S  I
ATGTCCGTTAAGCGATGTAAGCATGTTTTAACGGTATCTTATACCGTTAAAGATAGCATT   15840

 A  K  T  Y  N  V  D  S  E  K  I  S  V  I  Y  N  S  V  N  K
GCTAAAACTTATAATGTAGATAGTGAGAAAATATCAGTAATTTATAATAGTGTTAATAAA   15900

 S  D  F  I  Q  K  K  E  K  E  N  Y  F  L  A  V  G  A  S  W
TCTGATTTTATACAAAAAAAAGAAAAAGAGAATTACTTTTTAGCTGTTGGTGCAAGTTGG   15960

 P  H  K  N  I  H  S  F  I  K  N  K  K  V  W  S  D  S  Y  N
CCACATAAAAATATTCATTCATTCATAAAAAATAAAAAAGTTTGGTCTGACTCTTATAAT   16020

 L  I  I  V  C  G  R  T  D  Y  A  M  S  L  Q  Q  M  V  V  D
TTAATTATTGTATGTGGTCGTACTGACTATGCAATGTCTCTCCAACAAATGGTCGTTGAT   16080

 L  E  L  K  D  K  V  T  F  L  H  E  V  S  F  N  E  L  K  I
CTGGAACTAAAAGATAAAGTGACTTTTTTACATGAAGTCTCATTTAATGAATTAAAGATT   16140

 L  Y  S  K  A  Y  A  L  V  Y  P  S  I  D  E  G  F  G  I  P
TTATATTCTAAAGCCTACGCGCTTGTTTATCCATCTATTGATGAGGGTTTTGGTATACCT   16200

 P  I  E  A  M  A  S  N  T  P  V  I  V  S  D  I  P  V  P  H
CCTATTGAAGCGATGGCATCAAATACTCCAGTTATAGTGTCCGATATACCAGTATTTCAT   16260

 E  V  L  T  N  G  A  L  Y  V  N  P  D  D  E  K  S  W  Q  S
GAAGTGTTAACCAATGGTGCATTATATGTGAATCCGGATGATGAAAAAAGCTGGCAGAGT   16320

 A  I  K  N  I  E  Q  L  P  D  A  I  S  R  F  N  N  Y  V  A
GCAATTAAAAATATAGAGCAGTTGCCTGATGCAATTTCCCGATTTAACAACTATGTCGCA   16380

                                                 End of wbaU
 R  Y  D  F  D  N  M  K  Q  M  V  G  N  W  L  A  E  S  K  *
CGGTATGACTTTGATAATATGAAGCAGATGGTTGGCAATTGGTTGGCGGAATCAAAA TAA  16440

 Start of wbaN
 M  K  I  T  L  I  I  P  T  Y  N  A  G  S  L  W  P  N  V  L
ATGAAAATAACATTAATTATTCCCACATATAATGCAGGGTCGCTTTGGCCTAATGTTCTG   16500

 D  A  I  K  Q  Q  T  I  Y  P  D  K  L  I  V  I  D  S  G  S
GATGCGATTAAGCAGCAAACTATATATCCGGATAAATTGATTGTTATAGACTCAGGTTCT   16560

 K  D  E  T  V  P  L  A  S  D  L  K  N  I  S  I  F  N  I  D
AAAGATGAAACGGTTCCGTTAGCCTCAGACCTGAAAAATATATCAATATTTAATATTGAC   16620

 S  K  D  F  N  H  G  G  T  R  N  L  A  V  A  K  T  L  D  A
TCTAAAGATTTTAATCATGGAGGAACCAGAAATTTAGCAGTTGCAAAAACTCTGGACGCT   16680
```

Figure 10/12

```
 D  V  I  I  F  L  T  Q  D  A  I  L  A  D  S  D  A  I  K  N
GATGTTATAATTTTTCTAACGCAAGATGCAATTCTCGCGGATTCGGATGCAATTAAAAAT  16740

 L  V  Y  Y  F  S  D  P  L  I  A  A  V  C  G  R  Q  L  P  H
TTGGTTTATTATTTTTCAGATCCATTGATAGCAGCGGTTTGTGGTAGACAACTTCCTCAT  16800

 K  D  A  N  P  L  A  V  H  A  R  N  F  N  Y  S  S  K  S  I
AAAGATGCTAATCCCCTTGCAGTGCATGCCAGAAATTTAATTATAGTTCAAAATCTATT  16860

 V  K  S  K  A  D  I  E  K  L  G  I  K  T  V  F  M  S  N  S
GTTAAAAGTAAGGCAGATATAGAAAAATTGGGTATTAAAACTGTATTTATGTCCAATTCT  16920

 F  A  A  Y  R  R  S  V  F  E  E  L  S  G  F  P  E  H  T  I
TTTGCTGCCTATCGCCGTTCCGTTTTTGAAGAGTTAAGTGGGTTTCCTGAACATACAATT  16980

 L  A  E  D  M  F  M  A  A  K  M  I  Q  A  G  Y  K  V  A  Y
CTTGCCGAGGATATGTTTATGGCGGCTAAGATGATTCAGGCGGGTTATAAGGTCGCCTAC  17040

 C  A  E  A  V  V  R  H  S  H  N  Y  T  P  R  E  E  F  Q  R
TGCGCTGAAGCGGTGGTAAGACACTCCCATAATTATACCCCGCGAGAAGAGTTTCAACGA  17100

 Y  F  D  T  G  V  F  H  A  C  S  P  W  I  Q  R  D  F  G  G
TATTTTGATACTGGTGTATTTCATGCTTGTTCTCCGTGGATTCAGCGTGACTTTGGCGGA  17160

 A  G  G  E  G  F  R  F  V  K  S  E  I  Q  F  L  L  K  N  A
GCCGGTGGTGAGGGTTTCCGCTTCGTAAAATCAGAGATTCAATTCCTGCTTAAAAATGCA  17220

 P  F  W  I  P  R  A  L  L  T  T  F  A  K  F  L  G  Y  K  L
CCGTTCTGGATTCCAAGAGCTTTATTAACAACCTTTGCTAAATTCTTGGGTTACAAATTA  17280

 G  K  H  W  Q  S  L  P  L  S  T  C  R  Y  F  S  M  Y  K  S
GGCAAGCATTGGCAATCTTTACCGTTGTCTACATGTCGCTATTTTAGCATGTACAAGAGT  17340

                    End of wbaN  Start of manC
 Y  W  N  N  I  Q  Y  S  S  S  K  E  I  K  *  M  S  F  L  P
TATTGGAATAATATCCAATATTCTTCGTCAAAAGAGATAAAA TAAATGTCTTTTCTTCCC  17400

 V  I  M  A  G  G  T  G  S  R  L  W  P  L  S  R  E  Y  H  P
GTAATTATGGCTGGCGGCACAGGTAGCCGTTTATGGCCGCTTTCACGCGAATATCATCCG  17460

 K  Q  F  L  S  V  E  G  K  L  S  M  L  Q  N  T  I  K  R  L
AAGCAGTTTCTAAGCGTTGAAGGTAAACTATCAATGCTGCAAAATACTATAAAGCGATTA  17520

 A  S  L  S  T  E  E  P  V  V  I  C  N  D  R  H  R  F  L  V
GCTTCACTTTCTACAGAAGAACCCGTTGTCATTTGCAATGACAGACACCGTTTCTTAGTC  17580

 A  E  Q  L  R  E  I  D  K  L  A  N  N  I  I  L  E  P  V  G
GCTGAACAACTCCGTGAAATTGACAAGTTAGCAAATAATATTATTCTCGAACCGGTAGGC  17640

 R  N  T  A  P  A  I  A  L  A  A  F  C  A  L  Q  N  A  D  N
CGTAATACTGCACCAGCGATCGCTCTTGCCGCGTTTTGTGCGCTCCAGAATGCTGATAAT  17700

 A  D  P  L  L  L  V  L  A  A  D  H  V  I  Q  D  E  I  A  F
GCTGATCCTCTTTTGTTGGTTCTTGCTGCAGATCATGTGATTCAGGATGAAATAGCTTTT  17760

 T  K  A  V  R  H  A  E  E  Y  A  A  N  G  K  L  V  T  F  G
ACGAAAGCTGTCAGACATGCTGAAGAATACGCTGCAAATGGTAAGCTTGTAACTTTTGGT  17820

 I  V  P  T  H  A  E  T  G  Y  G  Y  I  R  R  G  E  L  I  G
ATTGTTCCAACGCATGCTGAAACGGGTTATGGATATATTCGTCGTGGTGAGTTGATAGGA  17880

 N  D  A  Y  A  V  A  E  F  V  E  K  P  D  I  D  T  A  G  D
AATGACGCTTATGCAGTGGCTGAATTTGTGGAGAAACCGGATATCGATACCGCCGGTGAC  17940

 Y  F  K  S  G  K  Y  Y  W  N  S  G  M  F  L  F  R  A  S  S
TATTTCAAATCAGGGAAATATTACTGGAATAGCGGTATGTTTTTATTTCGTGCAAGCTCT  18000
```

Figure 10/13

```
 Y  L  N  E  L  K  Y  L  S  P  E  I  Y  K  A  C  E  K  A  V
TATTTAAACGAATTAAAGTATTTATCACCTGAAATTTATAAAGCTTGTGAAAAGGCGGTA  18060

 G  H  I  N  P  D  L  D  F  I  R  I  D  K  E  E  F  M  S  C
GGACATATAAATCCCGATCTTGATTTTATTCGTATTGATAAAGAAGAGTTTATGTCATGC  18120

 P  S  D  S  I  D  Y  A  V  M  E  H  T  Q  H  A  V  V  I  P
CCGAGTGATTCTATCGATTATGCAGTTATGGAGCACACACAGCATGCGGTGGTGATACCA  18180

 M  S  A  G  W  S  D  V  G  S  W  S  S  L  W  D  I  S  N  K
ATGAGCGCTGGCTGGTCGGATGTGGGTTCCTGGTCCTCACTTTGGGATATATCGAATAAA  18240

 D  H  Q  R  N  V  L  K  G  D  I  F  A  H  A  C  N  D  N  Y
GATCATCAGAGAAATGTTTTAAAAGGAGATATTTTCGCACATGCTTGTAATGATAATTAC  18300

 I  Y  S  E  D  M  F  I  S  A  I  G  V  S  N  L  V  I  V  Q
ATTTATTCCGAAGATATGTTTATAAGTGCGATTGGTGTAAGCAATCTTGTCATTGTTCAA  18360

 T  T  D  A  L  L  V  A  N  K  D  T  V  Q  D  V  K  K  I  V
ACAACAGACGCTTTACTGGTGGCTAATAAAGATACAGTACAAGATGTTAAAAAAATTGTC  18420

 D  Y  L  K  R  N  D  R  N  E  Y  K  Q  H  Q  E  V  F  R  P
GATTATTTAAAACGGAATGATAGGAACGAATATAAACAACATCAAGAAGTTTTCCGCCCC  18480

 W  G  K  Y  N  V  I  D  S  G  K  N  Y  L  V  R  C  I  T  V
TGGGGAAAATATAATGTGATTGATAGCGGCAAAAATTACCTCGTTCGATGTATCACTGTT  18540

 K  P  G  E  K  F  V  A  Q  M  H  H  H  R  A  E  H  W  I  V
AAGCCGGGTGAGAAATTTGTGGCGCAGATGCATCACCACCGGGCTGAGCATTGGATAGTA  18600

 L  S  G  T  A  R  V  T  K  G  E  Q  T  Y  M  V  S  E  N  E
TTATCCGGGACTGCTCGTGTTACAAAGGGAGAGCAGACTTATATGGTTCTGAAAATGAA  18660

 S  T  F  I  P  P  N  T  I  H  A  L  E  N  P  G  M  T  P  L
TCAACATTTATTCCTCCGAATACTATTCACGCGCTGGAAAATCCTGGAATGACCCCCCTG  18720

 K  L  I  E  I  Q  S  G  T  Y  L  G  E  D  D  I  I  R  L  E
AAGTTAATTGAGATTCAATCAGGTACCTATCTTGGTGAGGATGATATTATTCGTTTAGAA  18780

                    Start of manB    End of manC
                                M  N  V  V  N  N  S  R  D  V
 Q  R  S  G  F  S  K  E  W  T  N  E  R  S  *
CAACGTTCTGGATTTTCGAAGGAGTGGACTAATGAACGTAGTTAATAATAGCCGTGATGT  18840

 I  Y  S  S  G  I  V  F  G  T  S  G  A  R  G  L  V  K  D  F
TATTTATTCATCAGGTATTGTGTTTGGAACGAGTGGGGCTCGCGGTCTTGTAAAAGATTT  18900

 T  P  Q  V  C  A  A  F  T  V  S  F  V  A  V  M  Q  E  H  F
TACACCTCAGGTATGTGCTGCTTTTACGGTTTCATTTGTTGCCGTTATGCAGGAACATTT  18960

 S  F  D  T  V  A  L  A  I  D  N  R  P  S  S  Y  G  M  A  Q
TTCCTTTGATACCGTAGCATTGGCAATAGATAATCGTCCAAGTAGTTATGGGATGGCTCA  19020

 A  C  A  A  A  L  A  D  K  G  V  N  C  I  F  Y  G  V  V  P
GGCGTGTGCTGCTGCATTGGCGGATAAAGGCGTTAACTGTATTTTTTATGGAGTGGTACC  19080

 T  P  A  L  A  F  Q  S  M  S  D  N  M  P  A  I  M  V  T  G
AACCCCAGCTTTGGCCTTTCAGTCTATGTCTGACAATATGCCTGCGATAATGGTTACGGG  19140

 S  H  I  P  F  E  R  N  G  L  K  F  Y  R  P  D  G  E  I  T
AAGTCATATTCCATTCGAGCGGAACGGCCTCAAGTTTTATCGTCCTGATGGTGAAATCAC  19200

 K  H  D  E  A  A  I  L  S  V  E  D  T  C  S  H  L  E  L  K
GAAACATGATGAGGCTGCGATCCTTAGTGTTGAAGATACGTGCAGCCATTTAGAGCTTAA  19260
```

Figure 10/14

```
  E  L  I  V  S  E  M  A  A  V  N  Y  I  S  R  Y  T  S  L  F
AGAACTCATAGTTTCAGAAATGGCTGCTGTTAATTATATATCTCGTTATACATCTTTATT  19320

  S  T  P  F  L  K  N  K  R  I  G  I  Y  E  H  S  S  A  G  R
TTCTACTCCATTCCTGAAAAATAAGCGTATTGGTATTTACGAACATTCAAGCGCTGGGCG  19380

  D  L  Y  K  P  L  F  I  A  L  G  A  E  V  V  S  L  G  R  S
TGATCTTTATAAGCCTTTATTTATTGCATTGGGGGCTGAAGTCGTTAGCTTGGGTAGAAG  19440

  D  N  F  V  P  I  D  T  E  A  V  S  K  E  D  R  E  K  A  R
CGATAATTTTGTACCTATAGATACAGAGGCTGTAAGCAAAGAGGATCGGGAAAAAGCTCG  19500

  S  W  A  K  E  F  D  L  D  A  I  F  S  T  D  G  D  G  D  R
CTCATGGGCTAAAGAGTTCGATTTAGATGCCATATTCTCGACAGATGGGGATGGTGATCG  19560

  P  L  I  A  D  E  A  G  E  W  L  R  G  D  I  L  G  L  L  C
CCCTCTTATTGCTGATGAGGCCGGTGAGTGGCTAAGAGGCGATATACTAGGTCTATTATG  19620

  S  L  A  L  D  A  E  A  V  A  I  P  V  S  C  N  S  I  I  S
TTCACTTGCATTGGATGCAGAAGCCGTCGCTATTCCTGTTAGTTGTAACAGCATAATTTC  19680

  S  G  R  F  F  K  H  V  K  L  T  K  I  G  S  P  Y  V  I  E
TTCTGGCCGCTTTTTTAAACATGTTAAGCTTACAAAAATTGGCTCGCCTTATGTTATCGA  19740

  A  F  N  E  L  S  R  S  Y  S  R  I  V  G  F  E  A  N  G  G
AGCTTTTAATGAATTATCGCGGAGTTATAGTCGTATTGTCGGTTTTGAAGCCAATGGCGG  19800

  F  L  L  G  S  D  I  C  I  N  E  Q  N  L  H  A  L  P  T  R
TTTTTTATTAGGAAGCGACATCTGTATTAACGAGCAGAATCTTCATGCCTTACCAACTCG  19860

  D  A  V  L  P  A  I  M  L  L  Y  K  S  R  N  T  S  I  S  A
TGATGCTGTATTACCAGCAATAATGCTGCTTTACAAAAGTAGGAATACCAGCATTAGCGC  19920

  L  V  N  E  L  P  T  R  Y  T  H  S  D  R  L  Q  G  I  T  T
TTTAGTCAATGAACTCCCAACTCGTTACACCCATTCTGACAGATTACAGGGGATTACAAC  19980

  D  K  S  Q  S  L  I  S  M  G  R  E  N  L  S  N  L  L  S  Y
TGATAAAAGTCAATCCTTAATTAGTATGGGCAGAGAAAATCTGAGCAACCTCTTAAGCTA  20040

  I  G  L  E  N  E  G  A  I  S  T  D  M  T  D  G  M  R  I  T
TATTGGTTTGGAGAATGAAGGTGCAATTTCTACAGATATGACAGATGGTATGCGAATTAC  20100

  L  R  D  G  C  I  V  H  L  R  A  S  G  N  A  P  E  L  R  C
TTTACGTGATGGATGTATTGTGCATTTGCGCGCTTCTGGTAATGCACCTGAGTTACGCTG  20160

  Y  A  E  A  N  L  L  N  R  A  Q  D  L  V  N  T  T  L  A  N
CTATGCAGAAGCTAATTTATTAAATAGGGCTCAGGATCTTGTAAATACAACGCTTGCTAA  20220

                    End of manB
  I  K  K  R  C  L  L  *
TATTAAAAAACGATGCTTGCTG TAAAAAAATTGAATGTTATTTACTTAATATGCCTATTT  20280

                                Start of wbaP
                                   M  D  N  I  D  N  K  Y
TATTTACATTATGCACGGTCAGAGGGTGAGGATTAAATGGATAATATTGATAATAAGTAT  20340

  N  P  Q  L  C  K  I  F  L  A  I  S  D  L  I  F  F  N  L  A
AATCCACAGCTATGTAAAATTTTTTTGGCTATATCGGATTTGATTTTTTTTAATTTAGCC  20400

  L  W  F  S  L  G  C  V  Y  F  I  F  D  Q  V  Q  R  F  I  P
TTATGGTTTTCATTAGGATGTGTCTATTTTATTTTTGATCAAGTACAGCGATTTATTCCT  20460

  Q  D  Q  L  D  T  R  V  I  T  H  F  I  L  S  V  V  C  V  G
CAAGACCAATTAGATACAAGAGTTATTACGCATTTTATTTTGTCAGTAGTATGTGTCGGT  20520
```


Figure 10/15

```
 W  F  W  I  R  L  R  H  Y  T  I  R  K  P  F  W  Y  E  L  K
TGGTTTTGGATTCGTTTGCGACATTATACTATCCGCAAGCCATTTTGGTATGAGTTAAAA  20580

 E  I  F  R  T  I  V  I  F  A  I  F  D  L  A  L  I  A  F  T
GAAATTTTTCGTACGATCGTTATTTTTGCTATATTTGATTTGGCTCTGATAGCGTTTACA  20640

 K  W  Q  F  S  R  Y  V  W  V  F  C  W  T  F  A  L  I  L  V
AAATGGCAGTTTTCACGCTATGTCTGGGTGTTTTGTTGGACTTTTGCCCTAATCCTGGTG  20700

 P  F  F  R  A  L  T  K  H  L  L  N  K  L  G  I  W  K  K  K
CCTTTTTTTCGCGCACTTACAAAGCATTTATTGAACAAGCTAGGTATCTGGAAGAAAAAA  20760

 T  I  I  L  G  S  G  Q  N  A  R  G  A  Y  S  A  L  Q  S  E
ACTATCATCCTGGGGAGCGGACAGAATGCTCGTGGTGCATATTCTGCGCTGCAAAGTGAG  20820

 E  M  M  G  F  D  V  I  A  F  F  D  T  D  A  S  D  A  E  I
GAGATGATGGGGTTTGATGTTATCGCTTTTTTTGATACGGATGCGTCAGATGCTGAAATA  20880

 N  M  L  P  V  I  K  D  T  E  I  I  W  D  L  N  R  T  G  D
AATATGTTGCCGGTGATAAAGGATACTGAGATTATTTGGGATTTAAATCGTACAGGTGAT  20940

 V  H  Y  I  L  A  Y  E  Y  T  E  L  E  K  T  H  F  W  L  R
GTCCATTATATCCTTGCTTATGAATACACCGAGTTGGAGAAAACACATTTTTGGCTACGT  21000

 E  L  S  K  H  H  C  R  S  V  T  V  V  P  S  F  R  G  L  P
GAACTTTCAAAACATCATTGTCGTTCTGTTACTGTAGTCCCCTCGTTTAGAGGATTGCCA  21060

 L  Y  N  T  D  M  S  F  I  F  S  H  E  V  M  L  L  R  I  Q
TTATATAATACTGATATGTCTTTTATCTTTAGCCATGAAGTTATGTTATTAAGGATACAA  21120

 N  N  L  A  K  R  S  S  R  F  L  K  R  T  F  D  I  V  C  S
AATAACTTGGCTAAAAGGTCGTCCCGTTTTCTCAAACGGACATTTGATATTGTTTGTTCA  21180

 I  M  I  L  I  I  A  S  P  L  M  I  Y  L  W  Y  K  V  T  R
ATAATGATTCTTATAATTGCATCACCACTTATGATTTATCTGTGGTATAAAGTTACTCGA  21240

 D  G  G  P  A  I  Y  G  H  Q  R  V  G  R  H  G  K  L  F  P
GATGGTGGTCCGGCTATTTATGGTCACCAGCGAGTAGGTCGGCATGGAAAACTTTTTCCA  21300

 C  Y  K  F  R  S  M  V  M  N  S  Q  E  V  L  K  E  L  L  A
TGCTACAAATTTCGTTCTATGGTTATGAATTCTCAAGAGGTACTAAAAGAACTTTTGGCT  21360

 N  D  P  I  A  R  A  E  W  E  K  D  F  K  L  K  N  D  P  R
AACGATCCTATTGCCAGGGCTGAATGGGAGAAAGATTTTAAACTGAAAAATGATCCTCGA  21420

 I  T  A  V  G  R  F  I  R  K  T  S  L  D  E  L  P  Q  L  F
ATCACAGCTGTAGGTCGATTTATACGTAAAACTAGCCTTGATGAGTTGCCACAACTTTTT  21480

 N  V  L  K  G  D  M  S  L  V  G  P  R  P  I  V  S  D  E  L
AATGTACTAAAAGGTGATATGAGCCTGGTTGGACCACGACCTATCGTTTCGGATGAACTG  21540

 E  R  Y  C  D  D  V  D  Y  Y  L  M  A  K  P  G  M  T  G  L
GAGCGTTATTGTGATGATGTTGATTATTATTTGATGGCAAAGCCGGGCATGACAGGTCTA  21600

 W  Q  V  S  G  R  N  D  V  D  Y  D  T  R  V  Y  F  D  S  W
TGGCAAGTGAGTGGGCGTAATGATGTTGATTATGACACTCGTGTTTATTTTGATTCCTGG  21660

 Y  V  K  N  W  T  L  W  N  D  I  A  I  L  F  K  T  A  K  V
TATGTTAAAAACTGGACGCTTTGGAATGATATTGCCATTCTGTTTAAAACAGCGAAAGTT  21720

                    End of wbaP
 V  L  R  R  D  G  A  Y  *
GTTTTGCGGCGAGATGGTGCGTATTAAGCTTACCGAGAAGTACTGAATAATAATTGTATA  21780

AATTAGCCTGCGTAAAATCTGAACGCATCAATCGCTACCTTAATATCATACCTTTGAGTT  21840
```


Figure 10/16

AACATACTATTCACCTTTAACCTGCCATGACCGTTTGTGGCAGGGTTTCCACACCTGACA 21900

GGAGTATGTAATGTCCAAGCAACAGATCGGCGTCGTCGGTATGGCAGTGATGGGGCGCAA 21960

CCTCGCGCTCAACATCGAAAGCCGTGGTTATACCGTCTCCGTTTTCAACCGCTCCCGTGA 22020

AAAGACCGAAGAAGTGATTGCCGAGAATCCCGGCAAAAAGCTGGTGCCTTATTACACGGT 22080

Figure 10/17

Please type a plus sign (+) inside this box → [+]

PTO/SB/01 (12-97)
Approved for use through 9/30/00. OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

DECLARATION FOR UTILITY OR DESIGN PATENT APPLICATION (37 CFR 1.63)

[X] Declaration Submitted with Initial Filing OR [ ] Declaration Submitted after Initial Filing (surcharge (37 CFR 1.16 (e)) required)

Attorney Docket Number	23541-01
First Named Inventor	Peter Richard REEVES
COMPLETE IF KNOWN	
Application Number	/ to be assigned
Filing Date	to be assigned
Group Art Unit	
Examiner Name	

As a below named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

NUCLEIC ACID MOLECULES SPECIFIC FOR BACTERIAL ANTIGENS AND USES THEREFOR

(Title of the invention)

the specification of which

[X] is attached hereto

OR

[ ] was filed on (MM/DD/YYYY) [ ] as United States Application Number or PCT International Application Number [ ] and was amended on (MM/DD/YYYY) [ ] (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment specifically referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR 1.56.

I hereby claim foreign priority benefits under 35 U.S.C. 119(a)-(d) or 365(b) of any foreign application(s) for patent or inventor's certificate, or 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or of any PCT international application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application Number(s)	Country	Foreign Filing Date (MM/DD/YYYY)	Priority Not Claimed	Certified Copy Attached? YES	Certified Copy Attached? NO
PO 6545	AU	05/01/1997	[ ]	[ ]	[X]
PO 8162	AU	07/22/1997	[ ]	[ ]	[X]
			[ ]	[ ]	[ ]
			[ ]	[ ]	[ ]

[ ] Additional foreign application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto.

I hereby claim the benefit under 35 U.S.C. 119(e) of any United States provisional application(s) listed below.

Application Number(s)	Filing Date (MM/DD/YYYY)

[ ] Additional provisional application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto.

Please type a plus sign (+) inside this box → [+]

PTO/SB/01 (12-97)
Approved for use through 9/30/00. OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

DECLARATION — Utility or Design Patent Application

I hereby claim the benefit under 35 U.S.C. 120 of any United States application(s), or 365(c) of any PCT international application designating the United States of America, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT international application in the manner provided by the first paragraph of 35 U.S.C. 112, I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR 1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application.

U.S. Parent Application or PCT Parent Number	Parent Filing Date (MM/DD/YYYY)	Parent Patent Number (if applicable)
PCT/AU98/00315	05/01/1998	

☐ Additional U.S. or PCT international application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto.

As a named inventor, I hereby appoint the following registered practitioner(s) to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith: ☐ Customer Number _____ → *Place Customer Number Bar Code Label here*

OR

☒ Registered practitioner(s) name/registration number listed below

Name	Registration Number	Name	Registration Number
Michael I. Wolfson	24,750	Morey B. Wildes	36,968
William H. Dippert	26,723		
R. Lewis Gable	22,479		

☐ Additional registered practitioner(s) named on supplemental Registered Practitioner Information sheet PTO/SB/02C attached hereto.

Direct all correspondence to: ☐ Customer Number or Bar Code Label _____ *OR* ☒ Correspondence address below

Name	William H. Dippert						
Address	Cowan, Liebowitz & Latman, P.C.						
Address	1133 Avenue of the Americas						
City	New York			State	NY	ZIP	10036-6799
Country	USA	Telephone	(212) 790-9200			Fax	(212) 575-0671

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Name of Sole or First Inventor: ☐ A petition has been filed for this unsigned inventor

Given Name (first and middle if any)	Family Name or Surname
Peter Richard	REEVES

Inventor's Signature								Date	2/5/10
Residence: City	Glebe	State	NSW	Country	AU	Citizenship	AU		
Post Office Address	20 Mansfield Street								
Post Office Address									
City	Glebe	State		ZIP	NSW 2037	Country	Australia		

☒ Additional inventors are being named on the ___ supplemental Additional Inventor(s) sheet(s) PTO/SB/02A attached hereto

Please type a plus sign (+) inside this box → [+]

PTO/SB/02A (2-97)
Approved for use through 9/30/98. OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

DECLARATION

ADDITIONAL INVENTOR(S) Supplemental Sheet
Page 1 of 1

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor						
Given Name (first and middle [if any])			Family Name or Surname				
Lei			WANG				
Inventor's Signature						**Date**	25/10/99
Residence: City	North Ryde	**State**	NSW	**Country**	AU	**Citizenship**	AU
Post Office Address	8A Holt Street						
Post Office Address							
City	North Ryde	**State**		**ZIP**	NSW 2113	**Country**	Australia

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor						
Given Name (first and middle [if any])			Family Name or Surname				
Inventor's Signature						**Date**	
Residence: City		**State**		**Country**		**Citizenship**	
Post Office Address							
Post Office Address							
City		**State**		**ZIP**		**Country**	

Name of Additional Joint Inventor, if any:	☐ A petition has been filed for this unsigned inventor						
Given Name (first and middle [if any])			Family Name or Surname				
Inventor's Signature						**Date**	
Residence: City		**State**		**Country**		**Citizenship**	
Post Office Address							
Post Office Address							
City		**State**		**ZIP**		**Country**	